Marcel Hackbart



Die Lehrperson als Golem oder Galatea

Empirische Untersuchungen zu Kompetenzerwartungen von Lernenden an Lehrpersonen

OPEN ACCESS



Die Lehrperson als Golem oder Galatea

Marcel Hackbart

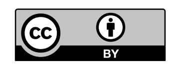
Die Lehrperson als Golem oder Galatea

Empirische Untersuchungen zu Kompetenzerwartungen von Lernenden an Lehrpersonen



Marcel Hackbart
Institut für Pädagogische Psychologie
Technische Universität Braunschweig
Braunschweig, Deutschland

Von der Fakultät für Geistes- und Erziehungswissenschaften der Technischen Universität Carolo-Wilhelmina zu Braunschweig zur Erlangung des Grades Doktorin / Doktor der Philosophie (Dr. phil.)* genehmigte Dissertation von Marcel Hackbart aus Röbel/Müritz.
Eingereicht am: 22.10.2021
Mündliche Prüfung am: 13.12.2021
Betreuer/in*: Prof. Dr. Barbara Thies
Weitere Referent/innen*: Prof. Dr. Elke Heise



ISBN 978-3-658-39443-1 ISBN 978-3-658-39444-8 (eBook)
https://doi.org/10.1007/978-3-658-39444-8

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Planung/Lektorat: Marija Kojic
Springer ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.
Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Inzwischen bearbeitete er mit glücklicher Hand und wundersamer Geschicklichkeit schneeweißes Elfenbein, gab ihm eine Gestalt, wie keine Frau auf Erden sie haben kann, und verliebte sich in sein eigenes Geschöpf.

(Ovid, 2003, in deutscher Übersetzung, S. 11, Original von −1/10)

Danksagung

Diese Arbeit konnte nur durch die Unterstützung und Begleitung vieler Menschen entstehen. Ihnen möchte ich an dieser Stelle großen Dank aussprechen.

Mein besonderer Dank gilt Prof. Dr. Barbara Thies, meiner Doktormutter, für die hervorragende Betreuung und enorme Unterstützung bei der Umsetzung der gesamten Arbeit – von der freien Auswahl des Themas über Hinweise zur Umsetzung des Promotionsprojekts bis zur Schaffung eines herausragenden Arbeitsumfelds. Danken möchte ich auch Prof. Dr. Elke Heise, meiner Zweitgutachterin, für ihr konstruktives Feedback und dafür, dass sie maßgeblich zum wunderbaren Arbeitsumfeld beigetragen hat.

An dieser Stelle möchte ich auch allen Mitarbeiter*innen, Praktikant*innen und Hilfskräften des Instituts für Pädagogische Psychologie der Technischen Universität Braunschweig, die mich direkt und indirekt unterstützt haben, herzlich danken. Ihre Hinweise, Rückmeldungen und Hilfestellungen sowie die außerordentlich freundschaftliche Kollegialität haben die Umsetzung des Promotionsprojekts ermöglicht und erleichtert. Auch den Kolleg*innen der Akademie Waldschlösschen gilt mein Dank für ihre Unterstützung in der finalen Phase der Fertigstellung.

Außerdem möchte ich mich bei allen Personen bedanken, die an den Untersuchungen teilgenommen bzw. zur Teilnahme aufgerufen haben. Ohne ihre Unterstützung wären die Studien und damit diese Arbeit nicht möglich gewesen.

Meinen tiefsten Dank möchte ich auch meinen Freund*innen sowie meinem Bruder und seiner Familie aussprechen. Die emotionale Unterstützung und die

manchmal auch notwendige Ablenkung waren ein wichtiger Baustein für die Umsetzung der Arbeit.

Schließend möchte ich mich bei einer ganz besonderen Person von ganzem Herzen bedanken, bei meiner Mutter Petra Hackbart. Ihr Durchhaltevermögen – auch in den größten Krisen mit wenigen Ressourcen – war für mich stets vorbildhaft. Ohne sie wäre ich nicht und wäre ich nicht dort, wo ich bin.

Zusammenfassung

Die Lehr-Lern-Interaktion ist ein aufeinander bezogenes Handeln von mindestens einer Lehrperson und einer lernenden Person (vgl. Hofer & Haimerl, 2008). Sie ist beschreibbar als ein komplexes und auch dynamisches Konstrukt (Knierim et al., 2017; Koopmans & Stamovlasis, 2016; Perrez et al., 2006; Thies, 2017), das mit verschiedenen Variablen in Lehr- und Unterrichtssituationen im Zusammenhang steht, z. B. Leistung, Motivation und sozial-emotionaler Kompetenz. In den Analysen von Hattie (2015) belegt die Lehr-Lern-Beziehung einen hohen Platz unter den bedeutsamsten Variablen für die Lernendenleistungen und -einstellungen in der Schule. Nickel (1985) postulierte in seinem transaktionalen Modell der Lehr-Lern-Interaktion, dass der soziokulturelle Kontext, in dem sich Menschen befinden, kognitive Schemata durch Sozialisations- und Erziehungserfahrungen bestimmt. Solche Schemata, wozu u. a. Erwartungshaltungen zählen, beeinflussen die Wahrnehmung der Unterrichtssituation sowie das Verhalten in ebendieser. Dies läuft zeitgleich bei Lehrpersonen und Lernenden ab, die in zirkulärer Wechselwirkung stehen. Eines der bekanntesten Phänomene, die die Lehr-Lern-Interaktion beeinflussen können, ist der Pygmalion-Effekt (Rosenthal & Jacobson, 1966, 1971), der die Auswirkungen von Lehrendenkompetenzerwartungen an Lernende beschreibt. Geht eine Lehrperson von einer kompetenten lernenden Person aus, verhält sie sich u. a. freundlicher, zugewandter und gibt mehr Gelegenheiten zur Kompetenzsteigerung (u. a. Harris & Rosenthal, 1985; Rosenthal, 1994). Dadurch bestätigen sich die Erwartungen der Lehrperson.

Der Pygmalion-Effekt ist vielfach untersucht. Seltener im Fokus stehen die Kompetenzerwartungen der Lernenden an die Lehrpersonen. Nichtsdestotrotz zeigte sich eine positive Wirkung hoher Lernendenkompetenzerwartungen an

Lehrpersonen auf die Lernenden selbst, z. B. hinsichtlich Lernleistung oder Verhalten (u. a. Jamieson et al., 1987). Wenig ist jedoch über die diesbezügliche Wirkung auf die Lehrpersonen bekannt. Die wenigen Befunde sind dabei teilweise widersprüchlich bzw. uneindeutig (Feldman & Prohaska, 1979; Feldman & Theiss, 1982). Deswegen wurden in Anlehnung an das Modell von Nickel (1985) Lernendenkompetenzerwartungen schrittweise empirisch betrachtet.

Zuerst wurde die Bedeutung des soziokulturellen Bezugsrahmens für die Ausbildung von Kompetenzerwartungen für verschiedene Gruppen von Lehrkräften adressiert sowie klassische Geschlechts-Professions-Assoziationen (u. a. Nosek et al., 2009) untersucht. Weder in Studie 1 ($N = 145$) auf Basis von bild- und textbasierten Fallvignetten noch in Studie 2 ($N = 254$) mit ausschließlich textbasierten Fallvignetten konnten Geschlechts-Professions-Assoziationen für die Unterrichtsfächer Mathematik und Deutsch ausgemacht werden, wenn für die Wärme- und Attraktivitätsbeurteilung kontrolliert wird. Anschließend stand in der Studie 3 das nonverbale Verhalten in Abhängigkeit von Lernendenkompetenzerwartungen an Lehrpersonen im Fokus. In einer Laborstudie wurden systematisch die Kompetenzerwartungen an eine Lehrperson (in einem Video) manipuliert und das nonverbale Verhalten der Teilnehmenden ($N = 62$) auf Basis von Videoaufnahmen ausgewertet. Es konnten Effekte für den Blickkontakt ausgemacht werden: Bei hohen Kompetenzerwartungen hielten die Teilnehmenden kürzere Zeit den Blickkontakt und sie sahen häufiger weg. Außerdem stützten sie (marginal signifikant) ihren Kopf seltener ab. Schließlich wurde in der vierten Studie dem Zusammenhang des ersten Kompetenzeindrucks, den Lehrpersonen machten, mit der Beurteilung einer Lehrveranstaltungssitzung und der Lehrpersonen – sowohl von Lernenden ($N = 871$ Angaben; Mehrfachteilnahmen möglich) als auch von den Lehrpersonen selbst ($N = 41$) – nachgegangen. Es zeigte sich mehrebenenanalytisch (unter Kontrolle des Wärmeeindrucks und der Attraktivitätsbeurteilung), dass die Lernendenkompetenzerwartungen positiv mit motivationalen Variablen (Interesse und Langeweile) und negativ mit dem Erleben von Störungen und Störungsmissmanagement aufseiten der Lernenden auf Gruppenebene (*between*) zusammenhingen – die Individualebene (*within*) hatte kaum Bedeutung. Der Einfluss der Lernendenkompetenzerwartungen auf die Lehrpersonen war unsystematischer: Es zeigten sich positive Zusammenhänge mit dem Erleben der eigenen Kompetenz und von positiven Emotionen sowie ein negativer Zusammenhang mit dem Erleben des eigenen Störungsmissmanagements.

Zusammengefasst zeigte sich, dass Lernendenkompetenzerwartungen an Lehrpersonen insgesamt die Lehr-Lern-Interaktion beeinflussen können. Die Effektstärken sind jedoch zumeist eher klein, insbesondere bei der Wirkung auf die Lehrpersonen selbst. Die Ergebnisse können im engen Sinn im Hinblick auf

die Bedeutsamkeit der Gestaltung des ersten Eindrucks interpretiert werden. Folglich ist es für Lehrpersonen förderlich, u. a. durch Stimmmodulation einen kompetenten Eindruck zu vermitteln. Weitergehend sind die Befunde aber auch bedeutsam für die Gestaltung der Lehr-Lern-Interaktion insgesamt, z. B. im Hinblick auf Machtdynamiken. Schließend wird ein synergetisch-systemischer Blick (u. a. Haken & Schiepek, 2010) auf Kompetenzerwartungen in der Lehr-Lern-Interaktion geworfen. Erwartungen und kognitive Schemata von Lernenden sind wirksame Einflüsse in der Lehr-Lern-Interaktion (vgl. Nickel, 1985), es sind aber weitere Untersuchungen nötig, um die Effekte von Lernendenkompetenzerwartungen valide beschreiben zu können.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Teil I

EINLEITUNG

1 Die Lehr-Lern-Interaktion

Das Bildungssystem, darunter insbesondere Schulen und Hochschulen, übernimmt nach Fend (2009) vier gesellschaftliche Funktionen:

1. Enkulturationsfunktion: „Reproduktion grundlegender kultureller Fertigkeiten und kultureller Verständnisformen" (Fend, 2009, S. 49)
2. Qualifikationsfunktion: Vermittlung von Fertigkeiten und Kenntnissen, „die zur Ausübung ‚konkreter' Arbeit erforderlich sind" (S. 50)
3. Allokationsfunktion: „Aufgabe, die Verteilungen auf zukünftige Berufslaufbahnen und Berufe vorzunehmen" (S. 50)
4. Integrations- und Legitimationsfunktion: Schaffung einer Identität, „die die innere Kohäsion einer Gesellschaft mitbestimmt" (S. 50), und „Schaffung von Zustimmung zum politischen Regelsystem" (S. 50)

Diese und weitere Aufgaben (z. B. Hochschule und Schule als Arbeitgeberinnen) müssen in einem komplexen Zusammenspiel unterschiedlicher Systeme aus individuellen Bedürfnissen von Personen verschiedener Gruppen (u. a. Lehrpersonen, Lernende und Erziehungsberechtigte), institutionellen Rahmenbedingungen sowie gesellschaftlichen und kulturellen Vorstellungen geleistet werden. Eine Bildungseinrichtung ist ein komplexes, dynamisches System und „[…] umfaßt das Systemganze (die […] Schulklasse einschließlich Lehrer […]), die Systemelemente selbst (die Schüler, die Lehrer, ggfs. die Schulleiter, die Schulordnung, usw.) und die wechselseitigen Beziehungen zwischen diesen Elementen (z. B. Lehrer-Schüler-Beziehungen)" (Brunner & Huber, 1991, S. 386).

Für die Aufgabenerfüllung des Bildungssystems ist die Lehr-Lern-Interaktion und -Beziehung wesentlich und wichtig. Sie kann angelehnt an Hofer und Haimerl (2008; s. a. Rosemann, 1978) wie folgt bestimmt werden: Die Lehr-Lern-Interaktion ist ein aufeinander bezogenes Handeln von mindestens einer

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_1

lehrenden und mindestens einer lernenden Person[1]. Sie ist ein multidimensionales, komplexes und dynamisches Konstrukt (Knierim et al., 2017; Koopmans & Stamovlasis, 2016; Perrez et al., 2006; Thies, 2017) und steht in der Schule im Zusammenhang mit Leistung, Motivation, der sozial-emotionalen Kompetenz (vgl. Knierim et al., 2017; Quin, 2016) und trägt zur Persönlichkeitsentwicklung bei (Siebertz-Reckzeh & Hofmann, 2017). Auch im Hochschulbereich zeigten sich positive Effekte für eine gelungene, zugewandte Interaktionsgestaltung (u. a. Komarraju et al., 2010; Micari & Pazos, 2012).

In den Betrachtungen von Hattie (2015) rangiert die Lehr-Lern-Beziehung auf Platz 11 der bedeutsamsten Variablen für die Lernendenleistungen und -einstellungen im schulischen Kontext mit einer mittleren Effektstärke von $d = 0.72$. Bei Hattie (2015) wurde jedoch ausschließlich die Meta-Analyse von Cornelius-White (2007) zugrunde gelegt, die sich insbesondere den personenzentrierten Variablen nach Rogers (1957, 1981) aufseiten der Lehrpersonen widmete. Es zeigte sich, dass insbesondere Nondirektivität ($r = .35$), Empathie ($r = .32$), Warmherzigkeit ($r = .32$) sowie die Förderung abstrakten Denkens ($r = .29$) im Zusammenhang mit positiven Effekten auf der Lernendenseite stehen. Ein empathischer, wertschätzender Umgang mit Lernenden fördert also die kognitive Entwicklung, die Motivation und das Wohlbefinden. Allerdings wird hier nur eine unidirektionale und paradigmatisch eng gefasste Perspektive auf die Lehr-Lern-Interaktion betrachtet.

Die Komplexität der Lehr-Lern-Interaktion schlägt sich in unterschiedlichen theoretischen Ansätzen nieder. In der Forschung standen zunächst unidirektionale Modelle im Fokus. So gibt es auf der einen Seite Ansätze, die die lernende Person als Ausgangspunkt der Lehr-Lern-Interaktion annahmen (Neill, 1973, nach Thies, 2017; Nohl, 1933, nach Thies, 2017), andere fokussieren die Lehrperson (z. B. Erziehungsstile von Lewin et al., 1939; Erziehungsdimensionen von Tausch & Tausch, 1963, nach Thies, 2017). Später entstanden komplexere interaktionale und transaktionale Modelle, die reine Ursache-Wirkungs-Beziehungen infrage stellten (*aptitude-treatment interaction* von Cronbach & Snow, 1977, nach Münzer & Brünken, 2018; transaktionales Modell von Nickel, 1985). Außerdem wurden (weitere) Ansätze aus der Erziehungsstilforschung („Freiheit in Grenzen" von Schneewind & Böhmert, 2009a, 2009b, 2009c, bei Kiel et al., 2013), der Bindungstheorie, der Selbstbestimmungstheorie (vgl. Knierim et al., 2017) und der Psychotherapie (u. a. humanistische Ansätze bei Cornelius-White, 2007;

[1] Häufig wird auch der Beziehungsbegriff herangezogen, da aber Beziehung ohne Interaktion nicht möglich ist, wird deswegen hier der Interaktionsbegriff genutzt.

Tausch, 2017; psychoanalytische Ansätze bei Lühmann, 2019; Schaub, 2013) auf die Lehr-Lern-Interaktion übertragen.

Die Lehr-Lern-Interaktion kann auch aus der Perspektive von Angebots-Nutzungs-Modellen (z. B. Helmke, 2017; Reusser & Pauli, 2010), den derzeit prominentesten Wirkmodellen der Lehr-Lern-Forschung (Kohler & Wacker, 2013; Seidel, 2014), betrachtet werden. Kernaussage dieser Theorien ist, dass Lehre und Unterricht ein Angebot darstellen, das nur durch die entsprechende Nutzung zu positiven Wirkungen, d. h. zu Kompetenzaufbau und weiteren Entwicklungen (z. B. der Persönlichkeit), führen kann. Das Unterrichtsangebot ist kontextgebunden und wird sowohl von den Lehrpersonen als auch den Lernenden bereitgestellt. Wenngleich es zumeist nicht explizit in den Modellen Erwähnung findet, ist auch der Beziehungsaufbau in der Lehre bzw. im Unterricht als Angebot der beteiligten Interaktionspartner*innen[2] anzusehen, das genutzt werden muss.

Lehre und Unterricht – sowohl in der Schule, der Hochschule als auch in der Erwachsenenbildung – sind durch eine asymmetrische Dependenz (Thies, 2010) sowie höhere Machtmittel aufseiten der Lehrpersonen (Misamer, 2019; vgl. Steins, 2014) bestimmt. Dennoch sind Lernende nicht einfach als abhängige Variablen zu betrachten, die ausschließlich auf Lehrpersonen reagieren, sondern sie sind aktiver Bestandteil in der Gestaltung der Lehr-Lern-Interaktion (vgl. u. a. Nickel, 1985; Thies, 2010).

Die meisten Modelle und Theorien zur Lehr-Lern-Interaktion sind zumeist genuin für die schulische und/oder familiäre Interaktion formuliert, können aber weitestgehend auch auf den Hochschulbereich und die Erwachsenenbildung übertragen werden. Erwachsenenbildung lässt sich nur tendenziell vom Lernen in der Schule abgrenzen (Siebert, 1996, nach Klein, 2011; vgl. Helmke et al., 2008). „Unabhängig vom Alter kann man zur Beschreibung und Erklärung von Lernprozessen die gleichen wissenschaftlichen Konstrukte heranziehen" (Stern, 2006, nach Klein, 2011, S. 120). Für einen Überblick zur Lehr-Lern-Interaktionsforschung sei hier auf die Herausgeberwerke von Schweer (2017; zum historischen Abriss s. Thies, 2017) und Herrmann (2019) verwiesen.

[2] In der vorliegenden Arbeit wird das Gendersternchen (*) genutzt, um alle Geschlechtlichkeiten zu berücksichtigen (vgl. Diewald & Steinhauer, 2020; Kronschläger, 2020). In den empirischen Untersuchungen in den Fragebögen wurde mit Splitting (/) gearbeitet.

2 Ziel und Aufbau der Arbeit

Erwartungen sind Antizipationen und können an verschiedene Ziele gerichtet werden (Kapitel 3), z. B. an Situationen, an Gegenstände, aber auch an Personen, sowohl auf andere als auch auf sich selbst (Kapitel 4). Personenbezogene Erwartungen können sich u. a. aus Stereotypen und Vorurteilen ergeben, da sich soziale Gruppen vermeintlich in bestimmten Eigenschaften unterscheiden. Das *stereotype content model* (u. a. Fiske & Taylor, 2017; Fiske et al., 1999; Fiske et al., 2002) stellt insbesondere die Wärme bzw. Sympathie und die Kompetenz als relevante Faktoren für die Beurteilung von Mitgliedern sozialer Gruppen heraus (Kapitel 5).

Der Pygmalion-Effekt wurde im Lehr-Lern-Kontext erstmalig im sogenannten Oak School-Experiment von Rosenthal und Jacobson (1966, 1971) untersucht und stieß die Forschung zu Kompetenzerwartungen von Lehrpersonen an (Kapitel 6). Rosenthal und Jacobson (1966, 1971) manipulierten nach dem Zufallsprinzip die Kompetenzerwartungen von Lehrkräften an ihre Schüler*innen, die sich selbst erfüllten: Schüler*innen, von denen die Lehrkräfte dachten, sie würden im Laufe des Schuljahres besonders viele Kompetenzen und Fähigkeiten aufbauen, schnitten in einem Test am Ende eines Schuljahres tatsächlich besser ab. Inzwischen ist der Pygmalion-Effekt mehrfach überprüft und bestätigt worden (u. a. Jussim & Harber, 2005; Ludwig, 2018). Damit der Pygmalion-Effekt wirksam werden kann, ist es notwendig, dass Lernende das Verhalten der Lehrperson hinnehmen und diesem keinen Widerstand entgegensetzen (Brophy & Good, 1976; Tauber, 1997). Erwartungen sind auch im Fokus von Theorien der Lehr-Lern-Interaktion, z. B. im transaktionalen Modell von Nickel (1985), das sich explizit auf den Pygmalion-Effekt beruft (Kapitel 7), sowie im Interaktionsregulationsmodell von Rosemann (1978), der insbesondere normative Erwartungen betrachtet

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_2

(Kapitel 8). Viele Kompetenzerwartungen von Lehrpersonen sind in natürlichen Lehr-Lenr-Situationen aufgrund von verschiedenen Stereotypen verzerrt (Kapitel 9).

Weitergehend können auch Lernende Erwartungen an Lehrpersonen haben (Feldman & Prohaska, 1979; Kapitel 10). Diese speisen sich u. a. auch aus den Gruppenzugehörigkeiten der Lehrkräfte (z. B. Geschlechtsstereotype), durch Informationen von Geschwistern bzw. Freund*innen oder durch öffentliche Lehrbewertungen im Internet (z. B. in Deutschland spickmich.de im Zeitraum von 2007 bis 2014 oder in den USA RateMyTeachers.com und RateMyProfessors.com). Die Lernendenerwartungen (Kapitel 11) an Lehrpersonen beeinflussen das Erleben und Verhalten der Lernenden, darunter auch die Leistung. Die Frage nach der Wirkung von Lernendenkompetenzerwartungen auf Lehrpersonen steht jedoch selten im Fokus. Bis heute sind in der englisch- und deutschsprachigen Literatur nur zwei Studien zu finden, die die Auswirkungen auf die Lehrpersonen erfassten (Feldman & Prohaska, 1979; Feldman & Theiss, 1982), die auch zu unterschiedlichen Befunden kamen: Bei Feldman und Prohaska (1979) konnte sich ein Einfluss der Lernendenkompetenzerwartungen auf die Lehrpersonen mediiert über das Verhalten zeigen, bei Feldman und Theiss (1982) blieb ein Effekt aus.

Die vorliegende Arbeit widmet sich empirisch Lernendenkompetenzerwartungen an Lehrpersonen, einem vernachlässigten Aspekt in der Lehr-Lern-Interaktion (Kapitel 12). In zwei experimentellen Studien werden stereotypbasierte Erwartungsverzerrungen auf Basis von Geschlecht betrachtet (Kapitel 13 und Kapitel 14). Anschließend wird das Verhalten von Lernenden in Abhängigkeit ihrer Kompetenzerwartungen an Lehrpersonen experimentell untersucht (Kapitel 15). Eine vierte Studie fokussiert Korrelate von Lernendenkompetenzerwartungen im Erleben von Lehre bei sowohl Lernenden als auch Lehrpersonen in einer Felduntersuchung (Kapitel 16). Im Anschluss an die Präsentation der vier empirischen Untersuchungen werden die Ergebnisse im Gesamten in Bezug auf theoretische Verortung (Kapitel 17), Limitationen und Stärken (Kapitel 18) sowie praktische und theoretische Ableitungen diskutiert (Kapitel 19).

Teil II

ERWARTUNGSEFFEKTE



3 Erwartungen und Erwartungseffekte

Erwartungen werden im allgemeinen Sprachgebrauch als eine vorausschauende Vermutung, Annahme oder Hoffnung beschrieben (Duden, o. J.). Unter Erwartungen in (motivations-)psychologischen Theorien wird „[…] die wahrgenommene Chance, dass sich aus einer Situation ein bestimmter Zielzustand ergibt" (Heckhausen & Heckhausen, 2018, S. 122) verstanden. Allgemeiner formuliert es Rudolph (2009, S. 21): „Eine Erwartung ist die subjektive Vorwegnahme eines Ereignisses […]."

> Der Begriff „Erwartung" wird im Zusammenhang mit dem Erwartungseffekt im deskriptiven Sinne verstanden, d. h. als mentale Antizipation zukünftiger Ereignisse, von deren Eintreffen der Erwartende überzeugt ist. Neben dieser Wortbedeutung wird „Erwartung" (*expectancy*) [...] auch im übertragenen, *normativen* Sinn gebraucht, d. h. zur Äußerung eines Wunsches, einer Forderung oder einer Norm. Wenn aus Perspektive des Erwartungseffekts hohe Leistungserwartungen als wünschenswert angesehen werden, sind also nicht hohe Leistungs*forderungen* gemeint [...]. (Ludwig, 2018, S. 142)

Erwartungen sind durch kognitive Prozesse determiniert (u. a. Thompson & Suñol, 1995) und sie sind eine subjektive Größe, die sich durch den soziokulturellen Bezugsrahmen und die Lerngeschichte der einzelnen Person ergeben (Heckhausen & Heckhausen, 2018; Nickel, 1985). Michael et al. (2012) beschreiben, dass, obwohl Erwartungen auf viele Arten entstehen können, sie oft das Ergebnis von Suggestionen sind, die von anderen Menschen oder von der Umwelt kommen. Wichtig ist, dass Erwartungen explizierbar sein können oder aber unbewusst das eigene Verhalten und Denken steuern (Michael et al., 2012).

Der Erwartungseffekt steht für „[…] einen hochspezifischen Einfluss von Antizipationen auf das Antizipierte" (Ludwig, 2018, S. 141) und beschreibt das Eintreten der eigenen Erwartung. Neben dem Begriff des Erwartungseffekts, der

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_3

in der vorliegenden Arbeit genutzt wird, wird auch die Bezeichnung „(sich) selbsterfüllende Prophezeiung" (*self-fulfilling prophecy*) im Lehr-Lern-Kontext synonym genutzt (vgl. Ludwig, 2018; Tauber, 1997). Wenn ein bestimmtes Ereignis erwartet wird, setzt sich automatisch eine Kette aus Kognitionen und Verhalten in Gang, um dieses Ereignis auszulösen, und Ursachen dafür werden fehlattribuiert (vgl. Geis, 1993; Michael et al., 2012). Erwartungseffekte lassen sich demnach mit zwei unterschiedlichen Phänomenen beschreiben (vgl. u. a. Greitemeyer, 2020):

1. Fehlinterpretation der Wahrnehmung: Ankommende Informationen werden konform mit den eigenen Erwartungen aufgenommen, verarbeitet und behalten.
2. Ausrichtung des eigenen Verhaltens, was eine Beeinflussung des Verhaltens anderer in erwartungskongruente Richtung nach sich zieht.

Häufig wird der Erwartungseffekt ausschließlich für bisher nicht eingetroffene bzw. nicht zutreffende Erwartungen genutzt (z. B. Jussim & Harber, 2005; Lorenz, 2019), d. h., es liegen noch keine tatsächlichen (bzw. „objektiven") Informationen zur erwarteten Eigenschaft vor. Jedoch können Erwartungseffekte auch weiter gefasst werden: Ein Kind, das bisher gute schulische Leistung erzielte, wird dadurch höhere Kompetenzerwartungen bei den Lehrkräften auslösen, die wiederum die Interaktion beeinflussen können. „Eine solche Variante der sich selbst erfüllenden Prophezeiung wird als *self-maintaining prophecy* [(sich) selbst erhaltende Prophezeiung] bezeichnet: Sie führt zur Beibehaltung eines Leistungsstandards bzw. vergrößert die ohnehin schon vorhandenen Leistungsunterschiede zwischen Schülern" (Ludwig, 2018, S. 142).

Obwohl die Auswirkungen von Erwartungen sehr stark sein können, ergeben sie sich nicht immer. In einigen Fällen können sie sich sogar umkehren, beispielsweise wenn eine Erwartung im Konflikt mit der bisherigen Erfahrung steht (Michael et al., 2012). Erlebt eine Lehrkraft beispielsweise über Jahre hinweg eine bestimmte Schülerin im Unterricht als besonders engagiert und fähig, kann die Aussage und Erwartungsinduktion von vermeintlichen Expert*innen, dass diese Schülerin sich in der Zukunft verschlechtern wird, dazu führen, dass die Erwartung nicht selbsterfüllend wird, sondern dass die Lehrkraft diese Schülerin besonders unterstützt. Dies kann eine Erklärung dafür sein, dass die Erwartungsmanipulationen, die von Rosenthal und Jacobson (1966, 1971) durchgeführt wurden, nur in den Klassenstufen 1 und 2 einen Einfluss auf die Lehr-Lern-Interaktion und somit auf die Leistung der Lernenden hatten und in den höheren Klassenstufen nicht (s. Kapitel 6).

Für interpersonale Kompetenzerwartungseffekte innerhalb des Lehr-Lern-Kontexts hat sich die Bezeichnung Pygmalion-Effekt von Rosenthal und Jacobson (1971) durchgesetzt (Ludwig, 2018). Neben dem Pygmalion-Effekt (s. Kapitel 6) gibt es noch eine Reihe von weiteren Erwartungseffekten sowie Differenzierungen.

4 Klassifikation von Erwartungseffekten

Es können drei unterschiedliche Klassen von Erwartungseffekten unterschieden werden: (1.) selbstbezogene Erwartungseffekte, (2.) fremdbezogene Erwartungseffekte und (3.) Erwartungs-Erwartungseffekte (s. Tabelle 4.1). Für diese Klassifikation wurde die ursächliche Erwartung für eine selbsterfüllende Prophezeiung herangezogen, im Prozess können aber weitere Erwartungseffekte relevant sein (Beispiel Pygmalion-Effekt s. Kapitel 6). In „natürlichen“ Situationen, also außerhalb von experimentellen Untersuchungen, sind somit die Effekte nicht immer eindeutig zu unterscheiden und können je nach Interpunktion (vgl. Watzlawick et al., 2017) unterschiedlich interpretiert werden. Die drei Klassen sollen im Folgenden überblicksartig kurz mit den zugeordneten Effekten skizziert werden.

Tabelle 4.1 Klassifikation von Erwartungseffekten

Selbstbezogene Erwartungseffekte	Fremdbezogene Erwartungseffekte	Erwartungs-Erwartungseffekte
• Placebo-Effekt • Nocebo-Effekt • Selbstwirksamkeits-erwartungen	• Versuchsleitungs-Erwartungseffekt • Pygmalion-Effekt • Galatea-Effekt • Golem-Effekt • Jastrow-Effekt	• Andorra-Effekt • Versuchsteilnehmende-Erwartungseffekte

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_4

4.1 Selbstbezogene Erwartungseffekte

Selbstbezogene Erwartungseffekte beziehen sich auf Erwartungen, die auf sich selbst und die eigenen Reaktionen gerichtet sind. Hierzu zählen der Placebo- und der Nocebo-Effekt, aber auch die Effekte von Selbstwirksamkeitserwartungen.

Ein Placebo ist ein Medikament ohne Wirkstoffe. Bei vielen Medikamentenstudien erreicht ein Placebo ähnliche Wirkungen wie ein Medikament mit Wirkstoffen, wenn den Patient*innen suggeriert wird, der Placebo habe eine Wirkung (vgl. Hansen et al., 2017; Hoefert, 2010). Durch die positiven Erwartungen an ein Medikament kann die heilende Wirkung verstärkt bzw. bei einem Placebo überhaupt erst angeregt werden. Dieses Phänomen wird als Placebo-Effekt bezeichnet und geht über Medikamente hinaus. So sind Placebo-Effekte auch bei u. a. psychotherapeutischen Maßnahmen beobachtbar (Lambert & Kleinstäuber, 2016; Locher et al., 2016). Der Nocebo-Effekt hingegen beschreibt die Zunahme von erwarteten Nebenwirkungen durch medizinische und therapeutische, insbesondere medikamentöse, Behandlungen (vgl. Hansen et al., 2017; Hoefert, 2010). Placebo-Effekte, aber auch Nocebo-Effekte in der Medizin finden sich empirisch abgesichert für verschiedene Bereiche (u. a. Fibromyalgie, Chen et al., 2017; neuropathischer Schmerz, Cragg et al., 2016; Morbus Menière, Dimitriadis & Panagiotis, 2017; chronisch-entzündliche Darmerkrankungen, Ma et al., 2019; Osteoarthritis, Zhang et al., 2008). Auch die implizite Lernleistung kann durch Placebo- und Nocebo-Effekte beeinflusst sein. In einem Experiment von Colagiuri et al. (2011) sollten Proband*innen an einem Wattepad riechen. Eine Gruppe erhielt die Information, dass das Riechen die kognitive Leistung steigert (Placebo-Gruppe), bei der zweiten Gruppe wurde behauptet, das Riechen reduziert die kognitive Leistung (Nocebo-Gruppe), und einer dritten Gruppe wurde keine Information zum Riechen gegeben. Es zeigte sich, dass die Placebo-Gruppe am besten ein Muster in Aufgabenlösungen unbewusst lernte (implizites Lernen) und die Nocebo-Gruppe am schlechtesten. Kirsch (1985) beschreibt drei Prozesse für den Placebo- und den Nocebo-Effekt in seiner *response expectancy theory*. Bei den Effekten handelt es sich um Erwartungen an die eigene Reaktion. Diese Erwartungen führen direkt zu (1) einer erwartungskonformen Wahrnehmung der eigenen Reaktion, (2) teilweise zu einer Änderung physiologischer Korrelate und (3) zu einer erwartungskonformen Anpassung des Verhaltens.

Eine weitere Form der selbstbezogenen Erwartungen ist die Selbstwirksamkeitserwartung. Diese wird bestimmt als die Überzeugung bzw. Erwartung, ein Verhalten mit Erfolg ausführen zu können, das benötigt wird, um ein Ziel oder ein gewünschtes Ergebnis zu erreichen (Bandura, 1977a, 1977b).

Selbstwirksamkeitserwartungen werden seltener im Kontext von Erwartungseffekten diskutiert, können jedoch auch wie ein Placebo wirken (vgl. Eden & Zuk, 1995). So ist es möglich, dass Personen, die hohe Selbstwirksamkeitserwartungen haben, bei Aufgaben mehr Ressourcen investieren, um erfolgreich zu sein, als Personen mit vergleichbaren Leistungspotenzial, die aber geringe Selbstwirksamkeitserwartungen haben, und dadurch auch mehr Erfolge erzielen. Dies kann aus meta-analytischen Ergebnissen abgelesen werden: Selbstwirksamkeitserwartungen stehen im Zusammenhang mit Persistenz (als ein Marker für den Ressourceneinsatz; $r = .38$, Multon et al., 1991), akademischer Leistung ($r = .38$, Multon et al., 1991) und Arbeitsleistung ($r = .38$, Stajkovic & Luthans, 1998).

4.2 Fremdbezogene Erwartungseffekte

Fremdbezogene Erwartungseffekte beziehen sich auf Erwartungen gegenüber anderen Personen (interpersonale Erwartungen) oder Objekten (objektbezogene Erwartungen; z. B. Computer-Lernprogramme). Hierzu gehören der Pygmalion-Effekt, der Versuchsleitungs-Erwartungseffekt und der Jastrow-Effekt.

Mit dem Versuchsleitungs-Erwartungseffekt (*experimenter expectancy bias*) wird der Einfluss des Wissens über die Hypothesen – also die Erwartungen – der Versuchsleitung in experimentellen Studien beschrieben. Es handelt sich um eine (unbewusste) Beeinflussung der Proband*innen bzw. der Versuchsobjekte zur Unterstützung der Hypothesen (Rosenthal & Fode, 1973; vgl. Klein et al., 2012) bzw. eine zugunsten der Hypothesen verzerrte Beobachtung und Auswertung (Beobachtungseffekte). Für eine detaillierte Darstellung sei hier auf Rosenthal (1966, 2009) verwiesen.

Der Pygmalion-Effekt beschreibt die selbsterfüllende Prophezeiung von Lehrendenerwartungen an Lernende (Rosenthal & Jacobson, 1966, 1971), kann aber auch darüber hinaus auf andere Kontexte Anwendung finden. Hat beispielsweise eine Lehrperson eine hohe Kompetenzerwartung an eine lernende Person, wird die lernende Person mehr Wissen generieren bzw. Fähigkeiten aufbauen. Dies wird über das Verhalten der Lehrperson vermittelt, die sich aufgrund ihrer Erwartungen emotional unterstützender verhält und forderndere Arbeitsmaterial zur Verfügung stellt (detaillierte Beschreibung des Pygmalion-Effekts s. Kapitel 6). Der Pygmalion-Effekt lässt sich unterteilen in den Galatea- und den Golem-Effekt (Babad et al., 1982). Der Galatea-Effekt beschreibt die selbsterfüllenden Auswirkungen einer positiven Erwartung (z. B. eine lernende Person ist fleißig), während

der Golem-Effekt die selbsterfüllenden Auswirkungen einer negativen Erwartung (z. B. eine lernende Person ist unfähig) beschreibt.

Begrifflich geht der Galatea-Effekt auf die Pygmalion-Mythologie zurück: Der Künstler Pygmalion erschuf die Statue einer Frau, in der seine ästhetischen Erwartungen vollends aufgingen, und verliebte sich in sie. Er bat Venus/Aphrodite, die Göttin der Liebe, um eine Frau, die so ist wie seine Statue aus Elfenbein. Die Göttin antwortete, indem sie die Statue zum Leben erweckte. Dieser Mythos beschreibt, wie die (unbewussten) Erwartungen von Pygmalion die Statue einer für ihn perfekten Frau entstehen ließen. Der Name der Statue, an die die Erwartungen von Pygmalion gerichtet wurden, ist Galatea (Ovid, 2003, Original von −1/10). Dieser Mythos ist in seiner Grundlogik von Rosenthal und Jacobson (1971) auf die Lehr-Lern-Interaktion übertragen worden, um die Relevanz von Erwartungseffekten (vgl. Ludwig, 2018) zu beschreiben. Der Golem-Effekt hingegen bezieht sich auf jüdische Legenden eines Golems, eines aus Ton geformten Wesens (vgl. Grözinger, 2009). Der Golem wurde ursprünglich zum Schutz der jüdischen Bevölkerung erschaffen, der mit der Zeit jedoch immer gewalttätiger wurde und sich gegen Menschen wendete, sodass er zerstört werden musste. Die negativen Auswirkungen von selbsterfüllenden Prophezeiungen wurden deswegen von Babad et al. (1982) als Golem-Effekt beschrieben. Pygmalion ist also mythologisch eine Figur, die (unbewusst) eine Erwartung hat, während Galatea und der Golem Figuren sind, an die (mehr oder weniger) eine Erwartung gerichtet wird.

In der Lehr-Lern-Interaktion zeigten sich sowohl Galatea- als auch Golem-Effekte. So schnitten beispielsweise bei Reynolds (2007) Studierende mit ähnlichen Ausgangsvoraussetzungen in einem Management-Kurs signifikant besser ab, wenn die dozierende Person dachte, die Gruppe hätte gute Leistungen (Galatea-Bedingung), im Vergleich zu einer Kontrollgruppe und zu Studierenden in einer Golem-Bedingung. Weitergehend schnitten die Studierenden der Golem-Bedingung schlechter ab als die Kontrollgruppe. Kritisch bei Reynolds (2007) anzumerken ist jedoch, dass es nur drei Dozierende gab (also eine dozierende Person je Bedingung), sodass der Effekt u. U. auch im Stil der Dozierenden unabhängig der Erwartungsinduktion begründet sein kann. (Kompetenz-)Erwartungen können aber auch bei Lernenden bestehen oder sich auf Objekte beziehen. Dies zeigte sich beispielsweise beim Lernen mit Computerprogrammen: Studierende, die ein qualitativ hochwertiges Lernprogramm erwarteten (Galatea-Bedingung), lernten mehr als Studierende, die ein Programm mit einer nicht eindeutigen Qualität (Golem-Bedingung) erwarteten. Studierende, die keine Erwartungen an das Lernprogramm hatten, schnitten mit einem Testwert zwischen den beiden anderen Gruppen ab (Fries et al., 2006; vgl. Haimerl & Fries, 2010).

Der Jastrow-Effekt beschreibt den Effekt, sich an aufgabenbezogenen Ergebnis-Erwartungen mit einem Werkzeug (im weiten Sinne) zu orientieren (Jastrow, 1901; vgl. Olshansky, 2007; Rosenthal & Jacobson, 1971). Wird eine maximale Obergrenze an Leistung mit einem bestimmten Werkzeug erwartet, wird sich an dieser Grenze orientiert. Dies beschrieb Jastrow (1901) am Beispiel der Hollorith-Tabelliermaschine: Im Jahr 1890 wurde die Maschine im Zensus-Amt der USA eingeführt. Es wurde einer ersten Gruppe von Arbeitenden beschrieben, dass mit der Maschine pro Tag etwa 550 Karten gelocht werden sollten. Nach zwei Wochen konnten einige Arbeitende 500 bis 550 Karten lochen, später mit öffentlichen Rekordlisten einige sogar bis zu 1500. Die höheren Leistungen gingen gleichzeitig mit hohem Anstrengungserleben einher. Einer zweiten Gruppe von Arbeitenden wurde keine Zahl von zu bearbeitenden Karten pro Tag genannt. Einige Arbeitende dieser zweiten Gruppen waren bereits nach drei Tagen in der Lage, 500 Karten zu bearbeiten – nach einer Woche konnten dies fast alle. Später waren Leistungen bis zu 2230 Karten möglich. Die Arbeitenden der zweiten Gruppe litten jedoch nicht so wie die Arbeitenden der ersten Gruppe an Überbelastung. Die Erwartung an die zu leistende Aufgabe hat das Verhalten und die Reaktion auf die Aufgabenerledigung bestimmt. Der Jastrow-Effekt wird hier den fremdbezogenen Erwartungseffekten zugeordnet, weil die Erwartungen sich auf die Möglichkeiten im Umgang mit einem Werkzeug bzw. einer Maschine (z. B. die Hollorith-Tabelliermaschine) beziehen. So schreibt Jastrow (1901), dass davon ausgegangen wurde, dass die Bedienung der Hollorith-Tabelliermaschinen ein hohes Maß an Intelligenz benötigt. Wird allerdings ein Leistungspotenzial von Mitarbeitenden erwartet, ist dies analog zum Pygmalion-Effekt. Sollten Mitarbeitende jedoch bestimmte Leistungen erbringen, weil sie glauben, diese erbringen zu müssen, liegen Erwartungs-Erwartungen vor.

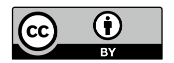
4.3 Erwartungs-Erwartungseffekte

Erwartungs-Erwartungseffekte beziehen sich auf angenommene Erwartungen anderer bezüglich der eigenen Person (vgl. antizipierte Fremdwahrnehmung, Thies, 2010). Hierzu zählt der Andorra-Effekt. Der Andorra-Effekt „[…] bezeichnet die Erscheinung, dass sich Menschen jenem Vorstellungsbild anpassen, das ihre Mitmenschen von ihnen haben. Wer für vertrauenswürdig gehalten wird, versucht das in ihn gesetzte Vertrauen zu rechtfertigen und ist selbst vertrauensfähig und vertrauensbereit“ (Nitsch 1975, zit. nach Lies et al., 2015, S. 171). Der Begriff Andorra-Effekt geht auf das Drama von Max Frisch (1999; uraufgeführt 1961) zurück, in dem der Protagonist als Jude ausgegeben und dadurch

negativen Stereotypen ausgesetzt ist. Mit der Zeit übernimmt er die negativen Verhaltensweisen, die ihm als Jude unterstellt werden, und wird später ermordet. Der Andorra-Effekt kann so erklärt werden, dass das vermittelte „Bild“ von außen in ein inneres Selbstbild integriert wird, das als Schema das eigene Verhalten (situationsspezifisch) beeinflusst (u. a. Froming et al., 1998; Pawlik & Buse, 1979). Parallelen für den Einfluss des Selbstbilds finden sich bei Sternzeichen. Es wird häufig diskutiert, ob Sternzeichen die Persönlichkeit von Menschen determinieren (vgl. Burke, 2012; Laux, 2008). Zumeist werden keine Zusammenhänge festgestellt und Befunde, die Unterschiede ausmachen, als Artefakte beschrieben (Hartmann et al., 2006; Wunder, 2003). Bei Pawlik und Buse (1979; Laux, 2008) zeigten sich hingegen bei Proband*innen, die an den Einfluss von Sternzeichen glaubten, Zusammenhänge zwischen Sternzeichen und Persönlichkeitseigenschaften. Der Glaube an Sternzeichen und Astrologie scheint moderierend zu wirken: Pawlik und Buse (1979; vgl. Laux, 2008) erklärten ihren Befund dadurch, dass die glaubenden Proband*innen das Wissen über Sternzeichen internalisiert hatten und dadurch ein (implizites) Selbstkonzept aufbauten, an das sich die Proband*innen anpassten. Fraglich bleibt, ob sich tatsächlich die Persönlichkeit dadurch verändert oder lediglich das Antwortverhalten (Laux, 2008).

Im Kontext der Erwartungseffekte werden z. B. auch der Hawthorne- und der John-Henry-Effekt erwähnt. Der Hawthorne-Effekt beschreibt die Veränderung des Verhaltens auf Basis einer Beobachtung bzw. einer Teilnahme an einer Studie (u. a. Wickström & Bendix, 2000) und der John-Henry-Effekt die Veränderung des Verhaltens (z. B. besondere Anstrengung) aufgrund der Annahme, zu einer Kontrollgruppe in einer experimentellen Studie zu gehören (u. a. Saretsky, 1972; auch kompensatorischer Wettstreit genannt, Döring & Bortz, 2016). Beide Effekte sind jedoch keine Erwartungseffekte im Sinne einer selbsterfüllenden Prophezeiung. Wenn Versuchsteilnehmende jedoch die Hypothesen, die überprüft werden sollen, kennen oder erahnen und sich diesen (bewusst oder unbewusst) anpassen, liegt ein Erwartungs-Erwartungseffekt vor (vgl. Döring & Bortz, 2016).

In dieser Arbeit sollen fremdbezogene, interpersonale Erwartungseffekte, die sich auf die Kompetenz anderer Personen beziehen, fokussiert werden. Kompetenzerwartungen können aus unterschiedlichen Quellen stammen: Mitteilung durch Dritte, öffentliche Bewertungen und Gerüchte, aber auch wesentlich durch Stereotype.


BY



5 Stereotypbasierte Erwartungen und das *stereotype content model*

> [The] real environment is altogether too big, too complex, and too fleeting, for direct acquaintance. We are not equipped to deal with so much subtlety, so much variety, so many permutations and combinations... We have to reconstruct it on a simpler model before we can manage with it. (Lippmann, 1922, nach Gilovich et al., 2006, S. 451)

Um die Komplexität der Welt zu reduzieren, greifen Menschen auf Schemata zurück. Ein Schema ist eine kognitive Repräsentation, die auf Kategorien aufbaut und vorverarbeitetes Wissen einer Person über Objekte, Personen oder Handlungen enthält (u. a. Fiske & Taylor, 2017; Gilovich et al., 2006; Pendry, 2014; Werth, Denzler & Mayer, 2020). Kategoriebasierte Verarbeitung ist im Vergleich zur individuierenden Verarbeitung weniger aufwendig, weniger anstrengend und schont kognitive Ressourcen (u. a. Fiske & Taylor, 2017; Klauer, 2020).

Stereotype sind Schemata, die sich auf soziale Gruppen und Mitglieder dieser Gruppen beziehen, und Resultat sozialer Kategorisierung (Fiske & Taylor, 2017; Greitemeyer, 2012; Klauer, 2020; Pendry, 2014; Werth, Denzler & Mayer, 2020). Grundlage für eine soziale Kategorisierung können verschiedene Merkmale wie Äußerliches (z. B. Hautfarbe, Geschlecht), Überzeugungen (z. B. Politik, Religion), Krankheiten oder Beeinträchtigungen (z. B. körperliche Behinderung, HIV) oder auch Verhalten (z. B. Kriminalität, Sexualität) sein (vgl. Klauer, 2020; Tröster & Pulz, 2020). Zu den bedeutsamsten Kategorien gehören die (zumeist sichtbaren) Merkmale Geschlecht, Ethnie und Alter, die auch als automatische bzw. primitive Kategorien bezeichnet werden (Fiske & Neuberg, 1990; Krings & Kluge, 2020).

Stangor (2009, S. 2) definiert Stereotype wie folgt und grenzt sie von Vorurteilen ab:

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_5

> We now define prejudice as a negative attitude toward a group or toward members of the group. Defining stereotyping has been more problematic—there are tens, if not hundreds of definitions in the literature, although they are mostly based on the general idea of stereotypes as knowledge structures that serve as mental "pictures" of the groups in question (Lippmann, 1922). With some exceptions, I'd say that we generally agree that stereotypes represent the traits that we view as characteristic of social groups, or of individual members of those groups, and particularly those that differentiate groups from each other. In short, they are the traits that come to mind quickly when we think about the groups.

Stereotype gehen also über die bloße Kategorisierung hinaus und beinhalten die Eigenschaften und die Erwartungen, die mit bestimmten sozialen Gruppen verknüpft sind (u. a. Klauer, 2020; Pendry, 2014; Werth, Denzler & Mayer, 2020). Dabei wird die Zugehörigkeit zu einer Gruppe herangezogen, um über ein Individuum ein Gesamturteil auf Basis eines Stereotyps zu erstellen (vgl. Halo-Effekt; vgl. Werth, Seibt & Mayer, 2020; Wolfrath, 2020) – unabhängig von der Korrektheit der spezifischen Kategorisierung. Stereotype stehen in Verbindung mit Vorurteilen und Diskriminierung. Dabei bezeichnen Stereotype eine kognitive Struktur, Vorurteile eine affektive und evaluative Struktur (in Kombination mit der kognitiven Struktur; Werth, Seibt & Mayer, 2020) sowie Diskriminierung einen konativen und verhaltensbezogenen Prozess (Gilovich et al., 2006; Kessler & Mummendey, 2007; Zanna & Rempel, 1988, nach Haddock & Maio, 2014).

Stereotype sind aufgrund unterschiedlicher Sozialisationsziele sowie -einflüsse (z. B. Medien) kulturspezifisch (Asbrock, 2010; Cuddy et al., 2009) und die Weitergabe von Stereotypen vollzieht sich durch Sozialisations- und Erziehungsprozesse (u. a. Cuddy et al., 2008). So zeigten sich Zusammenhänge in den Einstellungen von Kindern und ihren Eltern gegenüber Immigrant*innen und ethnischen Minderheiten (Rodrígez-García & Wagner, 2009) sowie für geschlechtsbezogene Stereotype (Tenenbaum & Leaper, 2002). Meta-analytisch ergab sich bei Degner und Dalege (2013) ein positiver Zusammenhang von Einstellungen gegenüber Fremdgruppen bei über 45000 Eltern-Kind-Paaren ($r = .29$). Die Entwicklung und der Aufbau von Stereotypen sind multikausal und können sowohl mit lerntheoretischen, (sozial-)kognitiven als auch motivationalen Ansätzen erklärt werden (Beelmann & Neudecker, 2020; Raabe & Beelmann, 2009). Stereotype werden aufgrund einer Reihe kognitiver Verzerrungen aufrechterhalten, die durch die limitierte Informationsverarbeitungskapazität entstehen (vgl. u. a. Fiske & Taylor, 2017; Gilovich et al., 2006; Petersen & Six, 2020; Spears & Tausch, 2014; Werth, Seibt & Mayer, 2020).

Das *stereotype content model* (u. a. Fiske & Taylor, 2017; Fiske et al., 1999; Fiske et al., 2002) macht Aussagen darüber, anhand welcher Kriterien soziale

Gruppen kategorisiert werden – und damit auch, welche Erwartungen an sie bestehen. Das Modell besagt, dass zwei Dimensionen von Bedeutung bei der Beurteilung von Stereotypen sind: Die Kompetenzwahrnehmung und die Wärmewahrnehmung, wobei die Wärmewahrnehmung der Kompetenzwahrnehmung im Modell theoretisch vorgelagert ist: Zuerst werden Menschen bzw. soziale Gruppen danach kategorisiert, ob sie kompetitiv und feindselig (= kalt) oder kooperierend und wohlwollend (= warm; Wärme in Anlehnung an den englischen Begriff *warmth* für Herzlichkeit/Freundlichkeit) sind. In einem zweiten Schritt wird der Status (= Kompetenz) beurteilt, d. h., ob sie in der Lage sind, kompetitiv zu handeln, oder potenziell eine Bedrohung sind. Dabei bewirkt die Kategorisierung, auf welche Weise gegenüber Personen bestimmter sozialer Gruppen reagiert wird (s. Tabelle 5.1). Sozial angesehene Menschen und Gruppen werden als warm (u. a. tolerant, warmherzig, ehrlich) und kompetent (u. a. unabhängig, intelligent, fähig) wahrgenommen.

Tabelle 5.1 Gruppenkategorien nach dem *stereotype content model* (nach Fiske et al., 2002)

	Kompetenz	
Wärme	Gering	Hoch
Hoch	Paternalistisches Vorurteil, z. B. älteren Menschen sowie „Hausfrauen" zugeschrieben: Mitglieder der Gruppen haben einen geringen Status und sind nicht kompetitiv. Sie lösen Mitleid und Mitgefühl aus.	Bewunderung, z. B. bei Mitgliedern der Eigengruppe und engen Verbündeten: Mitglieder haben einen hohen Status und sind nicht kompetitiv. Sie lösen Stolz und/oder Bewunderung aus.
Gering	Verachtendes Vorurteil, z. B. armen Menschen sowie Sozialleistungsempfangenden zugeschrieben: Mitglieder haben einen geringen Status und sind kompetitiv. Sie lösen Verachtung, Ekel und/oder Wut aus.	Neidisches Vorurteil, z. B. Asiat*innen, Jüd*innen und Feminist*innen zugeschrieben: Mitglieder haben einen hohen Status und sind kompetitiv. Sie lösen Neid und Missgunst aus.

Anmerkungen. Die Kategorisierung der Gruppen entspricht der stereotypen Zuschreibung, die nicht die Realität beschreibt.

Das zum *stereotype content model* analoge ABC-Modell der Stereotype (Koch et al., 2016) geht von drei Dimensionen aus: *agency*, *belief system* und *communion*. Das Modell postuliert ein zweidimensionales Muster, in das sich soziale Gruppen einordnen lassen. Diese beiden Dimensionen sind *agency*

(sozioökonomischer Erfolg) und *belief system* (Konservativität vs. Progressivität). Personengruppen, die durchschnittlich auf den beiden genannten Dimensionen abschneiden, bekommen hohe Werte bei *communion* (Gemeinschaftssinn). Dabei gibt es Parallelen zwischen *agency* im ABC-Modell und Kompetenz im *stereotype content model* sowie zwischen *communion* und Wärme.

Es bleibt festzuhalten, dass Stereotype neben den Eigenschaftszuschreibungen immer zeitgleich auch Erwartungen an Mitglieder von sozialen Gruppen bereitstellen. Diese sind in der Lehr-Lern-Interaktion besonders bedeutsam, wenn sich das Halten von Stereotypen in ungleichen Chancen und Lerngelegenheiten für Lernende auswirkt.

Teil III

ERWARTUNGEN IN DER LEHR-LERN-INTERAKTION

Check for updates

6 Der Pygmalion-Effekt

Der bereits erwähnte und kurz beschriebene Pygmalion-Effekt und damit die selbsterfüllenden Prophezeiungen im Lehr-Lern-Kontext wurden durch das Oak School-Experiment von Rosenthal und Jacobson (1966, 1971) bekannt und berühmt. Die sogenannte Oak School stand in einer größeren Stadt und die meisten Schüler*innen stammten aus der „unteren Klasse", wobei niemand aus besonders prekären Verhältnissen stammte. Im Mai 1964 wurden im Rahmen einer Studie der „Harvard National Science Foundation" mehr als 500 Kinder in Kindergartenklassen bis zur Klassenstufe 5 mit dem *test of general ability* (TOGA; bestehend aus einem Reasoning- und einem verbalen Subtest) überprüft (Vortest). Am Ende des Sommers 1964, vor Schuljahresbeginn, erhielten die Lehrkräfte eine Rückmeldung, welche Kinder ihrer Klasse auf Basis des Vortests *intellectual bloomers* (geistig Schnellentwickelnde bzw. Aufblühende) seien. Es wurden insgesamt ein Fünftel der Kinder als *intellectual bloomers* bezeichnet; die Auswahl beruhte aber nicht auf den Testergebnissen, sondern auf einer zufälligen Ziehung. „Der Unterschied zwischen den Kindern, die als intellektuell besonders entwicklungsfähig gekennzeichnet worden waren, und den Kontrollkindern, bei denen diese Kennzeichnung nicht stattgefunden hatte, bestand somit nur in den Köpfen der Lehrer" (Rosenthal & Jacobson, 1971, S. 93). Ein vorläufiger Wiederholungstest wurde im Januar 1965, der grundlegende Nachtest im Mai 1965 mit weniger als 400 Kindern und ein weiterer Wiederholungstest im Mai 1966 mit weniger als 300 Kindern (u. a. weil Kinder die Schule wechselten) ausgerichtet. Alle Tests wurden von Lehrkräften durchgeführt und von Forschungsassistent*innen, die blind gegenüber den Versuchsbedingungen waren, ausgewertet. Während nach einem Semester keine signifikanten Unterschiede zwischen Experimental- und Kontrollgruppe zu verzeichnen waren, gab es zum grundlegenden Nachtest, also nach einem Schuljahr, einen signifikanten Effekt (s. Abbildung 6.1): Schüler*innen der Experimentalgruppe erzielten

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_6

über alle Klassenstufen hinweg insgesamt vier Punkte mehr beim TOGA als Schüler*innen der Kontrollgruppe. Werden die einzelnen Klassenstufen getrennt betrachtet, ergaben sich nur für die Klassenstufen 1 ($\Delta M = 15$, $p < .05$) und 2 ($\Delta M = 10$, $p < .05$) signifikante Unterschiede. Erstaunlich ist, dass nach zwei Jahren (ein Jahr nach dem grundlegenden Nachtest) nur für die Klassenstufe 5 ein signifikanter Vorteil für die Experimentalgruppe ($\Delta M = 11$, $p < .05$) zu beobachten war.

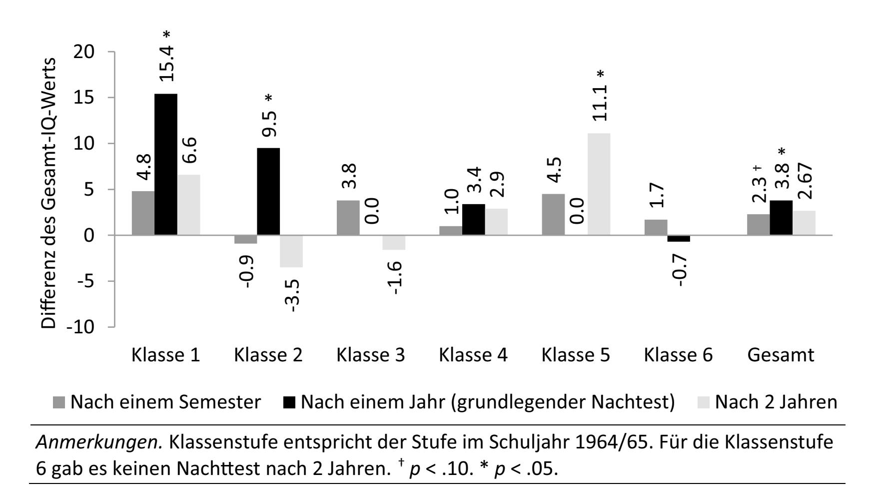


Abbildung 6.1 Erwartungsvorteil für *intellectual bloomers* beim Oak School-Experiment im Gesamt-IQ nach Zeit (Rosenthal & Jacobson, 1971)

Angelehnt an die griechische Mythologie (s. Kapitel 4), bezeichneten Rosenthal und Jacobson (1971) dieses Phänomen als Pygmalion-Effekt, da Pygmalion seine Erwartungen und Vorstellungen in die Statue hineinmeißelte – wie die Erwartungen der Lehrkräfte die Schüler*innen „formten“. Rosenthal und Jacobson (1971) erklärten diesen Erwartungseffekt wie folgt: „Vielleicht behandelten die Lehrer ihre Schüler auch freundlicher und liebevoller und zeigten ein größeres Interesse für sie, wenn sie positivere Erwartungen über deren intellektuelle Entwicklung hatten“ (S. 199). Brophy und Good (1976; vgl. Tauber, 1997) beschreiben den Pygmalion-Effekt in verschiedenen Phasen: Die Lehrperson entwickelt unterschiedliche (Kompetenz-)Erwartungen an die Lernenden und behandelt sie in Abhängigkeit der Erwartungsausprägung unterschiedlich. Das Verhalten der Lehrperson transportiert somit die Erwartung. Dadurch verändern

sich das Selbstkonzept und die Motivation der Lernenden sowie die Interaktion mit der Lehrperson, was die Erwartungen der Lehrperson verstärkt und – je nach Erwartungsausprägung – zu geringeren oder höheren Leistungen führt. Sind die Erwartungen der Lehrpersonen jedoch flexibel, wird sich mit der Zeit eine realistische, zutreffende Erwartung ausbilden.

Wie sich das Verhalten der Lehrpersonen auf Basis der Erwartungen genau verändert, beschreibt Rosenthal in einer 4-Faktoren-Theorie (u. a. Harris & Rosenthal, 1985; Rosenthal, 1994). Demnach werden Erwartungen über die Atmosphäre (*climate*), das Feedback, den Input und die Output-Möglichkeiten mediiert. Lehrpersonen schaffen für vermeintlich kompetente Lernende (wie im Fall der *intellectual bloomers*) ein wärmeres sozioemotionales Klima und sie geben schneller, kontingenter und differenzierter Feedback auf Antworten (vermeintlich) kompetenter Lernender. Weitergehend geben Lehrpersonen Lernenden, an die sie hohe Kompetenzerwartungen haben, mehr Input, d. h., diese Lernenden erhalten mehr Material und schwierigere Aufgaben, und mehr Output-Möglichkeiten, also mehr Möglichkeiten, sich zu äußern und sich in den Unterricht zu integrieren. Die Mediation über diese vier Faktoren konnte auch meta-analytisch untermauert werden (Harris & Rosenthal, 1985). Auch in einer jüngeren Untersuchung ergab sich bei Rubie-Davis (2007), dass Lehrkräfte mit hohen Kompetenzerwartungen an die Leseleistung der unterrichteten Klasse u. a. mehr Feedback gaben, mehr offene Fragen stellten, häufiger transparente Aufgabenstellungen gaben (*procedural statements*) und häufiger positives Verhaltensmanagement (im Sinne positiver Verstärkung) zeigten.

Die Ergebnisse von Rosenthal und Jacobson (1966, 1971) sind schon früh kritisiert worden (z. B. Elashoff & Snow, 1972; vgl. u. a. Jussim & Harber, 2005; Lorenz, 2019; Ludwig, 2018). Dass die Intelligenzentwicklung ungewöhnlich hoch und die Intelligenzmessung unreliabel sowie nicht objektiv war (z. B. haben teilweise die Lehrkräfte selbst die ersten Tests durchgeführt), stand allen Kritiken voran. Weitergehend zeigten sich die Unterschiede zwischen Experimental- und Kontrollgruppe nur in den unteren Jahrgängen (Klasse 1 und 2) und nicht mehr in den höheren (Klasse 3 bis 6), sodass insgesamt nur ein schwacher Effekt zu beobachten ist ($d = 0.28$, eigene Berechnung auf Basis der berichteten Mittelwerte, Standardabweichungen und Stichprobengrößen bei Rosenthal und Jacobson (1966, 1971); Raudenbush (1984) berichtet eine Gesamteffektstärke von $d = 0.21$). Außerdem hätten die Lehrkräfte durch die Cover-Story, es ginge um die Identifikation von Hochbegabten, ein besonderes Verpflichtungs- und Verantwortungsgefühl gegenüber den vermeintlichen *intellectual bloomers* spüren können. Auch konnte sich ein Großteil der Lehrkräfte in

anschließenden Interviews kaum an die vermeintlich hochbegabten Kinder erinnern. All diese Überlegungen führen zu der Frage, wie groß der Einfluss der Kompetenzerwartungen von Lehrpersonen ist.

Hattie (2015) kam unter Berücksichtigung von acht Meta-Analysen, in denen insgesamt 784 Effekte berücksichtigt wurden, auf eine Effektstärke von $d = 0.43$ („erwünschte Effekte") von Lehrendenkompetenzerwartungen für den Unterricht. Dies entspricht einem kleinen bis mittleren Effekt (Cohen, 1988). Bei einer genaueren Betrachtung der eingegangenen Meta-Analysen stellt sich heraus, dass diese selten den Effekt von Lehrendenkompetenzerwartungen auf Lernende fokussieren (wie beim Pygmalion-Effekt), sondern auch die Ausbildung bzw. Quellen von (Kompetenz-)Erwartungen betrachten. Auch gingen Kompetenzerwartungen über das Lernen bei Tieren in den von Hattie (2015) berichteten Effekt ein. Es ist also nicht eindeutig, was genau Hattie (2015) unter „Lehrererwartungen" fasst. Weitergehend ist stellenweise intransparent, wie Hattie (2015) die mittleren Effektstärkemaße bestimmte und wie die berichteten Angaben zustande kamen. So stimmt beispielsweise die Anzahl der eingegangenen Studien für die Meta-Analyse von Smith (1980) mit 47 nicht mit der Angabe von 46 bei Hattie (2015) überein. Außerdem ermittelte Raudenbush (1984) eine mittlere Effektstärke von $d = 0.11$, die Hattie (2015) mit $d = 0.08$ angibt.

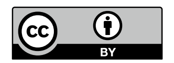
Smith (1980) berichtet in einer Meta-Analyse einen Einfluss der Lehrendenkompetenzerwartung auf die Lernendenleistung mit einer Effektstärke von Glass' $\Delta = 0.38$[1, 2] und auf den IQ mit einer Effektstärke von Glass' $\Delta = 0.16$. Raudenbush (1984) bestimmte eine ähnliche, mittlere Effektstärke von $d = 0.11$ für Lehrendenkompetenzerwartungen auf den IQ von Kindern. Einen Wert von $d = 0.54$ ermittelten Rosenthal und Rubin (1978) für die Kategorie Lernen und Fähigkeiten (darunter IQ-Testergebnisse und verbale Fähigkeiten) sowie einen Wert von $d = 0.88$ für alltägliche Situationen wie Symbollernen und athletische Leistung. Insgesamt machten Rosenthal und Rubin (1978) eine Effektstärke von $d = 0.70$ von Kompetenzerwartungen auf Leistungsergebnisse aus (über 345 Experimente, darunter auch Studien zum Lernen bei Tieren). Später bestimmte Rosenthal (1991, nach Rosenthal, 1994) meta-analytisch über 464 Studien eine Effektstärke von $d = 0.63$.

[1] Glass' Δ ist ein Effektstärkemaß, bei dem eine Mittelwertsdifferenz an der Standardabweichung einer Kontrollgruppe normiert wird (vgl. Bühner & Ziegler, 2009).

[2] Im Fließtext beschreibt Smith (1980), dass die Basis für die Effektstärke auf 44 Effekten beruht, in der Tabelle berichtet sie 22 eingegangene Effekte. An einigen Stellen stimmt die Anzahl der Effekte in der Hauptkategorie nicht mit der summierten Anzahl der Effekte in den Subkategorien überein.

Jussim und Harber (2005) stellen in einem Review heraus, dass es unklar ist, ob durch Erwartungseffekte der IQ beeinflusst werden kann. Insgesamt wird jedoch an anderen Stellen resümiert, dass die Effekte für akademische Leistung i. d. R. größer ausfallen als die für die allgemeine intellektuelle Fähigkeit (Smith, 1980; vgl. u. a. Lorenz, 2019, Ludwig, 2018). Trotz Kritik an der Originalstudie von Rosenthal und Jacobson (1966, 1971) halten Jussim und Harber (2005) fest, dass Erwartungseffekte in Lehr-Lern-Kontexten auftreten und der Pygmalion-Effekt besonders bei beginnenden Interaktionen sowie bei Interaktionen mit Lernenden aus stigmatisierten Gruppen bedeutsam ist.

Der Pygmalion-Effekt ist ein fremdbezogener Erwartungseffekt, aber im Prozess können auch selbstbezogene Erwartungen und Erwartungs-Erwartungen bedeutsam sein (s. Kapitel 4). So werden durch die Wahrnehmung des Verhaltens und die Äußerungen der Lehrpersonen durch die Lernenden das akademische Selbstbild der Lernenden beeinflusst (vgl. Brophy & Good, 1976) und weitergehend selbstbezogene Erwartungen induziert. Darüber hinaus (oder alternativ) können die Lernenden auch Erwartungs-Erwartungen aufbauen, z. B. bei geringen Lehrendenkompetenzerwartungen („Egal, was ich mache und wie ich mich anstrenge, die Lehrperson nimmt meine Anstrengungen nicht wahr, also lohnt sich die Mühe nicht“). In der Lehr-Lern-Interaktion können diese Prozesse zeitversetzt, aber auch parallel verlaufen. Das bedeutet, dass Lehrendenkompetenzerwartungen in die Interaktion hineingetragen werden (z. B. durch vorherige Informationen über Lernende oder Stereotype) und ursächlich für Interaktionsmuster (fremdbezogener Erwartungseffekt) sein können, aber sie können auch Resultat von selbstbezogenen Erwartungen und/oder Erwartungs-Erwartungen aufseiten der Lernenden sein.


BY

7 Das transaktionale Modell der Lehr-Lern-Beziehung

Der Pygmalion-Effekt und der Prozess der selbsterfüllenden Prophezeiung sind Phänomene der Lehr-Lern-Interaktion, die Bestandteile des transaktionalen Modells von Nickel (1985; Original von 1976; s. Abbildung 7.1) sind. Das transaktionale Modell der Lehr-Lern-Beziehung von Nickel (1985) erweitert unidirektionale Modelle der pädagogischen Beziehung und postuliert, dass Lernende und Lehrpersonen gleichermaßen den Interaktionsverlauf beeinflussen. Lernende und Lehrpersonen stehen in einer echten „Wechselbeziehung […], an der beide Seiten aktiven Anteil haben und durch die sie ihr Verhalten gegenseitig beeinflussen" (Nickel, 1985, S. 267), es finden also ständig Rückkopplungsprozesse statt. Nickel (1985) postuliert weiterhin, dass sowohl intrapsychische als auch soziokulturelle Phänomene die Lehr-Lern-Interaktion mitbestimmen.

Zu den intrapsychischen Bedingungsvariablen zählt Nickel (1985) Einstellungen, Erwartungshaltungen, darunter Rollenerwartungen, implizite Persönlichkeits- und Führungstheorien sowie verinnerlichte Werte, (Gruppen-)Normen und Gewohnheiten. Diese Variablen bilden ein „kognitives Modell" (Nickel, 1985, bezieht sich hier auf Argyle, 1972; vgl. relationale Schemata bei Thies, 2010), das zwischen Reizen und Verhalten vermittelt. Dieses kognitive Modell lässt einerseits das Verhalten anderer Personen beurteilen und „filtern", sodass es zu verzerrten Wahrnehmungen und Interpretationen des Verhaltens kommen kann, und andererseits bestimmt es das eigene Verhalten. So kann ein Selbstverständnis der Lehrperson als wissensvermittelnde und weniger als erziehende Person dazu führen, dass Verhaltensauffälligkeiten hauptsächlich in Misserfolgen (mangelnde Anstrengung und Begabung) begründet gesehen werden (vgl. Nickel, 1985). Außerdem sollte mit diesem Selbstverständnis das Unterrichtsverhalten der Lehrperson von einem eher an Erziehungsprozessen orientierten Verhalten der Lehrperson verschieden sein, was sich z. B. in einem eher fordernden und distanzierten Stil bemerkbar machen kann.

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_7

Die kognitiven Modelle bilden sich vor dem Hintergrund eines soziokulturellen Bezugsrahmens (vgl. sozialisationstheoretisches Modell des Unterrichts, Helmke, 2017). Dieser ist sowohl bei den Lehrpersonen als auch bei den Lernenden durch die soziale Lernvergangenheit (z. B. bisherige Erziehungserfahrungen), die gegenwärtigen sozialen Beziehungen (z. B. Familie, Mitschüler*innen bzw. Kolleg*innen) und durch objektivierte Einflüsse (z. B. Medien, Richtlinien) bestimmt (s. Abbildung 7.1). Zusätzlich entstehen Erwartungen und Einstellungen gegenüber einer spezifischen Person innerhalb der Interaktion (interaktionale Schemata bei Thies, 2010). So kann es sein, dass eine Professorin von einem Studenten antizipatorisch erwartet, er sei faul, weil sie diese Information über den Studenten von einem Kollegen erhalten hat. Allerdings zeigt sich der Student in der Lehrveranstaltung engagiert und bringt sich mit vielen qualitativ wertvollen Wortmeldungen ein. Dies kann die vorherige Erwartung umstrukturieren. Zusätzlich hat auch der Student Erwartungen an die Professorin, die parallel zu denen der Professorin in der Interaktion wirken. Mit der Zeit stellt sich dann ein bestimmtes Interaktionsmuster ein, das sich durch Rückkopplungsprozesse stabilisiert und durch Zirkularität (u. a. Haken & Schiepek, 2010; Hubrig & Herrmann, 2014; Strunk & Schiepek, 2006) gekennzeichnet ist, also durch eine gegenseitige Wechselwirkung.

Insbesondere der Pygmalion-Effekt (Rosenthal & Jacobson, 1966, 1971; s. Kapitel 6) wird von Nickel (1985) aufgegriffen, um sein Modell zu unterstützen und zu illustrieren. Die Information durch Forschende (also vom soziokulturellen Bezugsrahmen), dass eine*ein Schüler*in im kommenden Schuljahr eine besondere Kompetenzsteigerung zeigen wird (*intellectual bloomer*), induziert bei der Lehrperson eine Erwartung. Im realistischen Kontext können Kompetenzerwartungen beispielsweise durch Mitteilung bisheriger Noten entstehen oder durch Stereotype (s. Kapitel 5 und Kapitel 9). Die Erwartung führt bei der Lehrkraft zu einem positiveren, zugewandteren, wärmeren, unterstützenderen Verhalten gegenüber dieser*diesem Schüler*in im Vergleich zum Rest der Klasse (vgl. u. a. Brophy & Good, 1976). Im Sinne der 4-Faktoren-Theorie (Harris & Rosenthal, 1985; Rosenthal, 1994) sind die Atmosphäre (*climate*), das Feedback, die Input- und die Output-Möglichkeiten relevant, um die Kompetenzerwartungen (indirekt) auszudrücken. Die Lehrkraft macht folglich ein anderes Unterrichtsangebot (vgl. Angebots-Nutzungs-Modelle; u. a. Helmke, 2017; Reusser & Pauli, 2010). Dieses Verhalten wird von der lernenden Person wahrgenommen, führt zu einer konkreten Erwartung gegenüber der Lehrperson („Die Lehrkraft ist nett und hilfsbereit“) und somit zu einem Verhalten, das das Angebot der Lehrperson nutzt. Diese Angebotsnutzung wiederum führt zu einem Kompetenzaufbau, den die Lehrperson (z. B. in Leistungstests) wahrnimmt, sowie zu einer Bestätigung

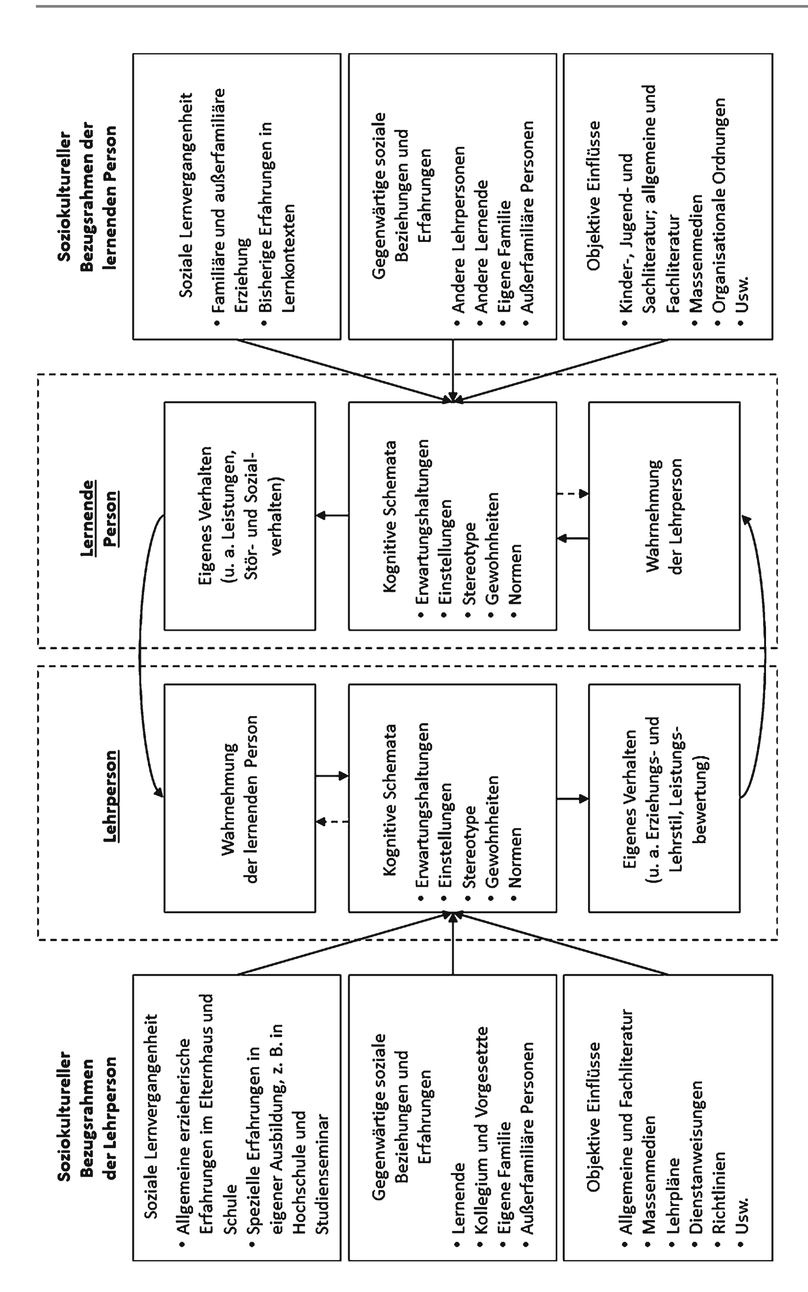


Abbildung 7.1 Transaktionales Modell der Lehr-Lern-Beziehung von Nickel (1985, S. 270)

der Erwartung und setzt einen „Engelskreislauf" (verstanden als positives Pendant zum Teufelskreislauf) in Gang. Im Laufe der Zeit entwickelt sich ein immer robusteres interaktionales (auf die dyadische Interaktion bezogenes) Schema (Thies, 2010), das im Sinne der selbsterhaltenden Prophezeiung wirken kann (Ludwig, 2018; s. Kapitel 3).



Normative Erwartungen in der Interaktionsregulation

8

Rosemann (1978) stellt neben den Erwartungen von Interaktionspartner*innen an die jeweils andere Person auch explizit die Bestätigung bzw. Widerlegung dieser Erwartungen in den Vordergrund seiner Überlegungen. Er spricht von Erwartungskonkordanz und -diskordanz. Die individuelle Erwartungskonkordanz ist dann gegeben, wenn eine Person A „[…] die Person B als in Übereinstimmung mit ihren an diese Person gerichteten Erwartungen erlebt. Bei Diskrepanzerlebnissen sprechen wir von Erwartungsdiskordanz“ (Rosemann, 1978, S. 42). Für sein Modell bezieht sich Rosemann (1978) auf normative Erwartungen, d. h. auf Vorstellungen und Schemata, wie ein Mensch in einer bestimmten Situation sein sollte und welche Rolle ihm zukommt. Es handelt sich folglich um ein Idealbild. Die Erwartungen und der Stellenwert der Erwartungen können zwischen unterschiedlichen Personen variieren und somit unterschiedliche Reaktionen auf das gleiche Verhalten hervorrufen. Als weitere Gruppe von Erwartungen diskutiert Rosemann (1978) Verhaltensantizipationen (antizipatorische Erwartungen; interaktionale Schemata bei Thies, 2010), die er als „[…] (gedankliche) Vorwegnahme des zukünftigen Verhaltens des [Interaktions-]Partners“ (S. 40) bestimmt und auf den Pygmalion-Effekt, in dem die Beschreibung der zukünftigen intellektuellen Entwicklung eines Kindes als Erwartungsinduktion dient (Rosenthal & Jacobson, 1966, 1971), bezieht (vgl. Schweer, 2019).

Lernende, unabhängig ob in Schule oder Hochschule, erwarten von Lehrpersonen i. d. R. fachliche Kompetenz, Freundlichkeit und Unterstützung sowie motivierenden Unterricht (u. a. Apel & Sandfuchs, 2003; Beishuizen et al., 2001; Denner et al., 2003; Merzyn, 2017; Richey, 2016; Sánchez et al., 2011; Yermack & Forsyth, 2016). Diese Erwartungen lassen sich in zwei unterschiedliche Themen einsortieren: in ein *love theme* (soziale Kompetenz) und ein *mastery theme* (Fachkompetenz; Ditton, 2002; Schweer & Rosemann, 1995). Dabei beinhaltet das *love theme* Wärme, Zuwendung sowie die Berücksichtigung von

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_8

Ängsten, Wünschen und Problemen der Lernenden. Fachwissen, qualitative Unterrichtsgestaltung, Disziplinierungs- und Durchsetzungsfähigkeit beschreiben das *mastery theme*. Lernende wünschen sich eher „demokratisch Führende" als Lehrpersonen (Denner et al., 2013), wobei die Relevanz der beiden Dimensionen zwischen Alter und Geschlecht variiert (Denner et al., 2013; Ditton, 2002).

Lehrkräfte erwarten hingegen angepasste, ordentliche und ruhige Schüler*innen (vgl. Richey, 2016; Rosemann, 1978): „Die Wunschschüler sind folgsame, interessierte, kreative, zur Selbsttätigkeit bereite Schüler, welche sich dem erzieherischen Handeln unterordnen, oder besser: es einsichtig anerkennen als Wert für ihre eigene Entwicklung, Bildung und Sozialisation" (Ntemiris, 2011, S. 92). Eigenschaften, die sich Dozierende von Studierenden wünschen, sind Teilnahme- und Durchhaltewille, Respekt und Anstand, Vorbereitung, Aufmerksamkeit, Reflexionsvermögen, Wissen und Kommunikationskompetenz (Voss & Eicher, 2015). Weitergehend zeigte sich, dass Dozierende als sehr wichtige Herausforderung für Studierende das Zeitmanagement, die Motivation für und die Fokussierung auf das Studium sowie die Aneignung einer wissenschaftlichen Arbeitshaltung benennen (Berthold et al., 2011), woraus sich auch Erwartungen an Studierende herausbilden. Im weiteren Sinne können auch hier das *love theme* (u. a. Respekt, Anstand, Angepasstheit im Sinne von Harmonieerhaltung) und das *mastery theme* (u. a. Motivation, Wille zum Aufbau von Wissen, Vorbereitung) wiedergefunden werden. Die normativen Erwartungen sowohl an Lernende als auch an Lehrpersonen lassen sich auch im Kontext des *stereotype content model* (s. Kapitel 5) betrachten: Das *love theme* entspricht weitestgehend der Wärme-Dimension und das *mastery theme* weitestgehend der Kompetenz-Dimension im *stereotype content model*.

Auf Basis dieser idealtypischen normativen Erwartungen wird nun das Verhalten einer konkret beobachteten Person wahrgenommen und bewertet. Dabei wird sich ein für die einzelne Person spezifisches Bild gemacht und dieses Bild führt dann zu gewissen Verhaltensantizipationen. Stimmen die Verhaltensantizipationen mit den normativen Erwartungen mehrheitlich überein, liegt also Erwartungskonkordanz vor, kommt es zu Zufriedenheit und folglich zu einem positiven Verhalten der wahrgenommenen Person gegenüber. Erwartungsdiskordanz hingegen führt nach Rosemann (1978) zu Unzufriedenheit und weniger positivem und/oder mehr negativem Interaktionsverhalten. Diese Prozesse laufen bei allen Interaktionspartner*innen ab und können über diverse Rückkopplungen zu positiven Engelskreisen oder negativen Teufelskreisen führen. Je nach normativen Erwartungen und erlebtem Verhalten wird ein anderes Angebot für die Lehr-Lern-Interaktion bereitgestellt (vgl. Angebots-Nutzungs-Modelle, z. B. Helmke, 2017; Reusser & Pauli, 2010).

Nimmt also beispielsweise eine Lehrperson eine Schülerin als ordentlich und ruhig wahr, baut sie ein interaktionales Schema und somit Verhaltensantizipationen auf („Diese Schülerin wird sich ordentlich und ruhig verhalten“). Sind diese mit den normativen Erwartungen konkordant, wird die Lehrperson wahrscheinlich zufrieden sein. Diese Zufriedenheit spiegelt sich dann im Verhalten wider, sodass sich die Lehrperson gegenüber der Schülerin positiver, d. h. unterstützender, wärmer und ermutigender verhält. Gleichzeitig nimmt die Schülerin die Lehrperson wahr, z. B. als kompetent und unterstützend. Dieses wahrgenommene Verhalten wird mit ihren normativen Erwartungen abgeglichen und bei Erwartungskonkordanz verhält sich die Schülerin leistungsbereiter, interessierter und weniger störend. Bei Erwartungskonkordanz von Studierenden wurden in einer Untersuchung von Schweer und Rosemann (1995) das Lehrverhalten, der Lernerfolg und das Vertrauen zur dozierenden Person signifikant günstiger eingeschätzt als bei Erwartungsdiskordanz.

Der wesentliche Unterschied zum transaktionalen Modell von Nickel (1985) liegt darin, dass sich das Verhalten einer Person nach Rosemann (1978) aus dem Abgleich von normativen Erwartungen (Idealvorstellungen) mit personenspezifischen Erwartungen (Verhaltensantizipation) einer*eines Interaktionspartner*in ergibt. Bei Nickel (1985) ergibt es sich direkt aus den (bereits vorweg bzw. mit dem ersten Eindruck entstandenen) positiven oder negativen Erwartungen gegenüber der*dem Interaktionspartner*in, die sich jedoch vergleichbar zu Rosemanns Ansatz durch personenspezifische Erfahrungen umstrukturieren können (vgl. Schweer, 2019).

Check for updates

9 Stereotypbasierte Lehrendenerwartungen

Kompetenzerwartungen von Lehrpersonen an Lernende können auf einer Reihe von Informationen beruhen, darunter auf Berichten von Kolleg*innen oder Erzählungen von Erziehungsberechtigten bzw. Angehörigen. Ohne solche Informationen spielen insbesondere Stereotype eine Rolle: Die stereotypen Bilder von Angehörigen bestimmter sozialer Gruppen gehen mit Eigenschaftszuschreibungen und somit auch mit Erwartungshaltungen einher (u. a. Stangor, 2009; s. Kapitel 5). In einer Meta-Analyse von Südkamp et al. (2012) konnte ein relativ hoher Zusammenhang zwischen der Lehrendeneinschätzung zur Leistung von Lernenden und der tatsächlichen akademischen Leistung der Lernenden ausgemacht werden ($r = .63$). Wenngleich Erwartungen häufig zutreffend sind, können sie dennoch stereotypbasiert zu selbsterhaltenden Prophezeiungen führen (vgl. Ludwig, 2018), sodass sich Lernende (im Sinne eines Andorra-Effekts) an die Lehrendenerwartungen anpassen (vgl. Lies et al., 2015; s. Kapitel 3 und Kapitel 4). Erste ggf. zutreffende Lehrendenerwartungen basierend auf Stereotypen (z. B. „Mädchen sind schlecht in Physik“) können also zuerst selbsterfüllend sein, dann aber selbsterhaltend werden, da sich die Lernenden innerhalb der Interaktion dem Stereotyp immer weiter annähern (z. B. „Ich bin ein Mädchen, ich kann kein Physik“), obwohl sie potenziell in der Lage wären, das Stereotyp zu widerlegen. Nach einem solchen Prozess kann die Lehrendeneinschätzung einen hohen Zusammenhang mit der akademischen Leistung haben, nicht aber mit dem akademischen Potenzial.

Bei der sozialen Wahrnehmung – auch in akademischen Settings – spielen in erster Linie die automatischen Kategorien (Fiske & Neuberg, 1990; Krings & Kluge, 2020) eine Rolle sowie die physische Attraktivität. Zu den automatischen Kategorien gehören Geschlecht, Ethnie und Alter, da sie sehr schnell (und automatisiert) über das optische Erscheinungsbild vermeintlich relativ zutreffende

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_9

Informationen liefern. Darüber hinaus beeinflusst auch der sozioökonomische Status der Lernenden die Lehrendenerwartungen (vgl. Brophy & Good, 1976). Im Folgenden wird ein Überblick über empirische Befunde zu stereotypbasierten Lehrendenerwartungen dargestellt.

9.1 Geschlecht[1]

In Alltagsdiskursen findet sich häufig die Auffassung, dass Mädchen und Frauen bessere sprachliche Fähigkeiten haben als Jungen und Männer. Letztere wiederum sind vermeintlich kompetenter im mathematischen und technischen Bereich (u. a. Nosek et al., 2009; Su et al., 2009). In Meta-Analysen konnte im Widerspruch zu solchen Stereotypen jedoch kein bedeutsamer Leistungsunterschied zwischen männlichen und weiblichen Lernenden in Mathematik-Leistungstests ($d = 0.05$; Lindberg et al., 2010) oder in verbalen Fähigkeiten ($d = 0.11$, dabei $d = 0.20$ für die generelle verbale Fähigkeit, Hyde & Linn, 1988; $d = 0.19$ für die Leseleistung, Lietz, 2006) ausgemacht werden. Dennoch wurden solche Stereotype von Lehrpersonen und angehenden Lehrpersonen teilweise berichtet (u. a. Li, 1999; Lorenz et al., 2016; Tiedemann, 1995, 2002). Auch bei Gentrup et al. (2018) hatten Grundschullehrkräfte signifikant höhere Erwartungen an Mädchen bezüglich ihrer Sprachfähigkeit und an Jungen hinsichtlich ihrer Mathematikleistung. Erwähnenswert ist zudem, dass die Lehrkräfte mehr Motivation und eine höhere Arbeitsmoral von den Mädchen im Vergleich zu den Jungen erwarteten.

Insgesamt zeichnet sich domänenunabhängig eine positivere Bewertung von weiblichen Lernenden im Vergleich zu männlichen ab und es wird von „männlichen Bildungsverlierern" (u. a. Diefenbach, 2010) gesprochen. Dies erscheint vor dem Hintergrund, dass die kommunalen Eigenschaften (z. B. kommunikativ, kooperativ, emotional), die eher Mädchen und Frauen zugeschrieben werden (u. a. Bosak et al., 2008; Diehl et al., 2004), dem Wunschbild einer lernenden Person (Ntemiris, 2011; Rosemann, 1978; Voss & Eicher, 2015; vgl. Kapitel 8) entsprechen, nicht verwunderlich. So schrieben Lehrkräfte in einer Studie von Heyder und Kessels (2015) insbesondere sich „maskulin" verhaltenden männlichen Schülern ein geringeres akademisches Engagement zu. Voyer und Voyer (2014) zeigten meta-analytisch, dass Frauen und Mädchen bessere Noten erzielen als ihre männlichen Kommilitonen bzw. Klassenkameraden ($d = 0.23$), wobei das

[1] Es sei hier darauf hingewiesen, dass es mehr als zwei Geschlechter gibt (vgl. Hackbart & Tischoff, 2020; Timmermanns & Böhm, 2020). Allerdings werden in der Forschung zu Stereotypen kaum Geschlechtlichkeiten über ein binäres Modell hinaus berücksichtigt (vgl. Hackbart, 2020).

Ausmaß über verschiedene Fächer variiert ($d = 0.37$ für Sprachen bzw. Sprachwissenschaften, $d = 0.07$ für Mathematik, $d = 0.15$ für Naturwissenschaften (*sciences*) und $d = 0.17$ für Sozialwissenschaften (*social sciences*)). Für den schulischen Kontext in Deutschland resümiert Budde (2008, S. 16): „In allen Fächern erhalten Jungen auch bei gleichen Kompetenzen schlechtere Noten.“ Wenngleich viele Studien keine bedeutsamen Unterschiede zwischen den Kompetenzerwartungen an Frauen bzw. Mädchen und Männern bzw. Jungen fanden (u. a. Auwarter & Aruguete, 2008), zeigten die meisten Studien mit Unterschieden leicht höhere Lehrendenkompetenzerwartungen an weibliche Lernende ($d = 0.20$ in der Meta-Analyse von Dusek & Joseph, 1983).

9.2 Ethnie und Migrationshintergrund[2]

In Deutschland leben etwa 20.1 Millionen Menschen mit Migrationshintergrund, das entspricht 24 % der Gesamtbevölkerung (im Kalenderjahr 2019; Statistisches Bundesamt, 2020a). Von ihnen hatten die meisten einen türkischen Migrationshintergrund oder eine türkische Migrationsbiografie (2.8 Millionen, 14 %; Statistisches Bundesamt, 2020a). Menschen mit türkischem Migrationshintergrund werden in Deutschland als prototypische Ausländer*innen angesehen (vgl. Cohrs & Asbrock, 2009) und ihnen werden weniger positive als negative Eigenschaften zugeschrieben (Froehlich et al., 2016; Kahraman & Knoblich, 2000). Auf den Dimensionen des *stereotype content model* werden Türk*innen kalt und inkompetent bis mittelhoch kompetent bewertet (Asbrock, 2010; Cuddy et al., 2007).

Glock und Krolak-Schwerdt (2013) ließen Lehramtsstudierende vermeintliche Kommiliton*innen auf der Basis textbasierter Fallvignetten bewerten, in denen die ethnische Zugehörigkeit mithilfe des Namens (Michael als deutscher, autochthoner Student oder Aslan als türkischstämmiger Student) und eine Kompetenzerwartung („Bester in der Klasse“ oder „unterdurchschnittlicher Student“) systematisch variierten. Insgesamt zeigte sich, dass der autochthone Student über alle Bedingungen hinweg als leistungsstärker bewertet wurde als der türkischstämmige Student. Allerdings ergaben sich bedeutsame Unterschiede nur bei der Golem-Bedingung („unterdurchschnittlicher Student“) für die verbale Leistungsfähigkeit. Für die mathematische Leistungsfähigkeit und in der

[2] Ethnie und Migrationshintergrund werden in dieser Arbeit synonym für nicht-autochthone Bevölkerungsgruppen verwendet.

Galatea-Bedingung konnten keine signifikanten Unterschiede in den Beurteilungen ausgemacht werden. Die Ergebnisse konnte Glock (2016) an einer Stichprobe mit berufstätigen Lehrkräften replizieren. Bei einem Vergleich von türkischstämmigen Kindern mit Kindern ohne Information zur Ethnie beurteilten Lehrkräfte und Lehramtsanwärter*innen die Leistung beschriebener Kinder ähnlich (Glock & Krolak-Schwerdt, 2014). Der eigene ethnische Background der Lehrperson scheint die Zusammenhänge jedoch zu moderieren. So zeigten Lehramtsanwärter*innen einer ethnischen Minderheit positivere Einstellungen gegenüber Schüler*innen mit Migrationsbiografie als Lehramtsanwärter*innen der ethnischen Majorität (Glock & Kleen, 2019).

Lorenz (2019) resümiert für Deutschland, dass Grundschullehrkräfte verzerrte Kompetenzerwartungen auf Basis des Migrationshintergrundes haben. So wurden türkischstämmige Schüler*innen sowohl im Fach Deutsch als auch in Mathematik als weniger kompetent angesehen als deutsche. Demgegenüber gab es keine Verzerrung bei Kindern mit osteuropäischem Migrationshintergrund im Fach Deutsch und eine Verzerrung der Kompetenzerwartungen zugunsten von Kindern mit osteuropäischem Migrationshintergrund im Fach Mathematik (vgl. Lorenz et al., 2016).

In anderen kulturellen Kontexten spielen aufgrund von Einwanderung andere Ethnien eine Rolle. So wird beispielsweise in Neuseeland zu Beginn eines Schuljahres ostasiatischstämmigen Kindern von Lehrkräften die höchste mathematische Kompetenz zugeschrieben und den Kindern der Mãori die geringste (Peterson et al., 2016). In der US-amerikanischen Meta-Analyse von Dusek und Joseph (1983) zeigte sich, dass „dunkelhäutigen" Menschen (*persons of color*; Afroamerikaner*innen und Mexikaner*innen) eine geringere Kompetenz zugeschrieben wird als „hellhäutigen". Dieser Effekt ist mit $d = 0.11$ jedoch sehr klein. In einer späteren Meta-Analyse von Tenenbaum und Ruck (2007) ergab sich das Bild, dass im Vergleich zu angloamerikanischen Lernenden asiatisch-amerikanische (*Asien Americans*) von Lehrpersonen als besonders kompetent ($d = 0.17$) und lateinamerikanische ($d = 0.46$) sowie afroamerikanische Lernende ($d = 0.25$) als weniger kompetent betrachtet werden.

9.3 Sozioökonomischer Status und Namen

Häufig sind Menschen, die einen Migrationshintergrund haben bzw. einer von der Mehrheit abweichenden Ethnie angehören, auch Mitglieder sozioökonomisch schwacher Gruppen. So ist der Anteil von Personen ohne Schulabschluss bei Menschen mit Migrationshintergrund (10 %) höher als bei autochthonen (1 %;

Statistisches Bundesamt, 2020a). Außerdem sind prozentual mehr Menschen mit Migrationshintergrund als autochthone erwerbslos (6 % zu 2 %) und haben im Durchschnitt ein geringeres Nettoeinkommen (1837 € zu 2225 €; Statistisches Bundesamt, 2020a). Aber auch Menschen ohne Migrationshintergrund können, wie die genannten Zahlen zeigen, einen geringen sozioökonomischen Status haben, der sich auf die interpersonalen Erwartungen auswirken kann.

Bei Auwarter und Aruguete (2008) zeigte sich, dass Lehrkräfte auf Basis von textbasierten Fallvignetten von Schüler*innen geringere Leistung in der Zukunft erwarten, wenn die Schüler*innen einen geringen sozioökonomischen Status haben. Glock und Krolak-Schwerdt (2014) konnten hingegen mithilfe von zwei Fallvignetten, die beide Kinder mit Migrationshintergrund beschrieben, keinen Einfluss des sozioökonomischen Status auf die Leistungsbewertung ausmachen. Bei Lehrkräften, die Lorenz et al. (2016) untersuchten, konnte ein Einfluss des sozioökonomischen Status auf die Lehrendenkompetenzerwartungen gefunden werden. Bei der Meta-Analyse von Dusek und Joseph (1983) werden mit einer Effektstärke von $d = 0.47$ Lernende aus dem Mittelstand für kompetenter gehalten als Lernende aus „unteren Schichten“.

Ist der sozioökonomische Status nicht bekannt, können auch Namen einen Hinweis liefern und Erwartungen formen (vgl. Dusek & Joseph, 1983). Bei Tobisch und Dresel (2017) erwarteten Lehrkräfte auf Basis des Namens von einem Murat (Migrationshintergrund und geringer sozioökonomischer Status) eine an den bisherigen Leistungsergebnissen angepasste Note, wohingegen von einem Justin (kein Migrationshintergrund und geringer sozioökonomischer Status) und von einem Julius (kein Migrationshintergrund und hoher sozioökonomischer Status) bessere Noten als bisher. Der sozioökonomische Status, der mit den Namen verknüpft wurde, hing signifikant mit der Einschätzung der generellen Fähigkeit und der der Anstrengung sowie mit Erwartungen an die Leistungen in Deutsch und in Sozialwissenschaften (*social sciences*) zusammen. Eine andere Umfrage zur Bewertung von Vornamen durch Lehrkräfte kam zu folgenden Ergebnissen (Kaiser, 2009, Absatz 1):

> Als eher freundlicher, leistungsstärker und verhaltensunauffällig stellen sich Kinder mit Vornamen wie Charlotte, Sophie, Marie, Hannah, Alexander, Maximilian, Simon, Lukas oder Jakob im Bewusstsein von LehrerInnen dar, während Namen wie Chantal, Mandy, Angelina, Kevin, Justin oder Maurice eher mit Leistungsschwäche und Verhaltensauffälligkeit assoziiert werden. Besonders „Kevin“ hat sich als stereotyper Vorname für einen „verhaltensauffälligen“ Schüler herausgestellt. In einem Fragebogen fand sich der Kommentar „Kevin ist kein Name, sondern eine Diagnose!“.

9.4 Physische Attraktivität

Physische Attraktivität bzw. Schönheit ist ein schwer zu definierendes Konstrukt und wird häufig als Mittel subjektiver Bewertungen operationalisiert (Langlois et al., 2000).

> Ein wichtiges Ergebnis [bisheriger Attraktivitätsforschung] ist beispielsweise die Erkenntnis, dass – anders als der Volksmund behauptet – Schönheit eben nicht (nur) im Auge des Betrachters liegt, sondern dass es bei allen „Geschmacksunterschieden" dennoch eine gemeinsame Basis für menschliche Attraktivität gibt, sowohl innerhalb einer Kultur als auch mit Abstrichen zwischen verschiedenen Kulturen. (Gründl, 2011, S. 1)

So werden beispielsweise symmetrische Gesichter mit reiner Haut als besonders attraktiv wahrgenommen.

Attraktiveren Menschen werden mehr positive Eigenschaften zugeschrieben als weniger attraktiven Menschen. Dies ist als *„beauty is good" stereotype* (bzw. *„what is beautiful is good" stereotype*; Dion et al., 1972) bekannt. Meta-analytisch zeigten Eagly et al. (1991), dass Attraktivität u. a. einen Effekt auf die Einschätzung der intellektuellen Kompetenz ($d = 0.46$ entspricht $r = .22$) hat. In einer späteren Meta-Analyse von Jackson et al. (1995) zeigte sich, dass der Zusammenhang von Attraktivität und der Wahrnehmung der intellektuellen Kompetenz bei $d = 0.65$ ($r = .31$) liegt. Dabei ist der Effekt am stärksten bei der Beurteilung von erwachsenen Männern ($d = 0.93$ entspricht $r = .42$), am zweitstärksten für erwachsene Frauen ($d = 0.70$ entspricht $r = .33$) und am schwächsten für Kinder ($d = 0.49$ entspricht $r = .24$). Auch bei Langlois et al. (2000) zeigten sich meta-analytisch Zusammenhänge zwischen Attraktivität und angenommener akademischer Kompetenz bei Kindern ($r = .48$) und der angenommenen beruflichen Kompetenz bei Erwachsenen ($r = .41$).

Das *„beauty is good" stereotype* zeigte sich auch in pädagogischen Kontexten: In einer viel beachteten Studie von Landy und Sigall (1974) wurde männlichen Studierenden entweder ein gutes oder ein schlechtes Essay vorgelegt. Zusätzlich wurde den Proband*innen ein Foto der vermeintlichen (weiblichen) Autorin vorgelegt, das entweder eine attraktive Person oder eine unattraktive zeigte. Eine weitere Gruppe erhielt keine Informationen über die Attraktivität der Verfasserin. Es zeigte sich, dass die guten Essays tatsächlich besser bewertet wurden. Zugleich zeigte sich aber auch, dass eine attraktive Verfasserin bessere Bewertungen erhielt als eine unattraktive Verfasserin oder eine Verfasserin mit unbekanntem Aussehen.

In einem Experiment von Shapiro et al. (2007) wurde das vermeintliche Gewicht als ein Merkmal von Attraktivität eines Trainees mithilfe von Bildern manipuliert, bevor ein Training via Gegensprechanlage durchgeführt wurde. Dabei zeigte sich, dass vermeintlich übergewichtigen Personen von den Lehrpersonen mit weniger hohen Erwartungen bezüglich des Trainingserfolgs und der Arbeitsmoral begegnet wurde. Außerdem beurteilten die Lehrpersonen das Training negativer und zudem weibliche Lehrpersonen die lernende Person negativer, wenn sie von einer übergewichtigen Person ausgingen. Müller et al. (2017) untersuchten in einem ähnlichen Design wie Glock und Krolak-Schwerdt (2013) bzw. Glock (2016) zu ethnischen Stereotypen die Bewertung übergewichtiger Studierender durch Lehramtsanwärter*innen. Dabei zeigte sich entgegen den Hypothesen der Autor*innen, dass ein beschriebener übergewichtiger Student in den Leistungs- und Kompetenzdomänen (Mathematik, niederländische Sprachfähigkeit, Intelligenz) konstant besser beurteilt wurde als ein beschriebener normalgewichtiger Student. Müller et al. (2017) erklären diesen Befund damit, dass entweder die Lehramtsanwärter*innen keine negativen Stereotype gegenüber übergewichtigen Menschen aufgebaut hatten oder dass sie bewusst stereotypes Denken kontrollierten. Insgesamt werden trotz einzelner, widersprechender Befunde jedoch meta-analytisch attraktive Lernende von Lehrpersonen akademisch fitter ($d = 0.30$ bei Dusek & Joseph, 1983; $d = 0.36$ bei Ritts et al., 1992) und sozial kompetenter ($d = 0.19$ bei Dusek & Joseph, 1983; $d = 0.48$ bei Ritts et al., 1992) bewertet als unattraktive Lernende.

9.5 Resümee zu stereotypbasierten Lehrendenerwartungen

Die Darstellung der empirischen Ergebnisse zu stereotypbasierten Lehrendenerwartungen konnte an dieser Stelle nur an exemplarischen Studien ausschnitthaft und reduziert dargestellt werden. So werden viele Zusammenhänge über verschiedene moderierende und mediierende Mechanismen erklärt. Beispielsweise zeigte sich in der Längsschnittstudie von Sorhagen (2013), dass im sprachlichen Bereich die unzutreffende Lehrendenkompetenzerwartung an Lernende in der ersten Klassenstufe die Leistungen in standardisierten Tests im Alter von 15 Jahren signifikant vorhersagen konnte. Allerdings wurde dieser Zusammenhang vom sozioökonomischen Status (Einkommen der Familie) moderiert, d. h., dass der Zusammenhang bei Kindern aus sozioökonomisch schwachen Familien größer war. Bei Kindern aus sozioökonomisch starken Familien war der Zusammenhang sehr klein oder kehrte sich bei einigen Tests sogar um, d. h., dass

Kinder, die in der ersten Klasse von ihren Lehrkräften unterschätzt wurden, im Alter von 15 Jahren besser abschnitten als Kinder, die überschätzt wurden. Für den mathematischen Bereich zeigten sich hingegen keine Langzeiteffekte.

Resümierend kann festgehalten werden, dass Leistungserwartungen auf Basis von Stereotypen existieren, die aber nicht bei jeder Lehrperson in jeder Situation zum Tragen kommen. In der Regel ist der Zusammenhang zwischen Gruppenzugehörigkeit der Lernenden und Leistungserwartungen der Lehrpersonen an die Lernenden eher klein (Geschlecht, ethnische Minoritätszugehörigkeit) bis moderat (Attraktivität, sozioökonomischer Status, ethnische Minoritätszugehörigkeit).



10 Die Lehrperson als Golem oder Galatea – Quellen der Erwartungen

Das transaktionale Modell von Nickel (1985) beschreibt die Lehr-Lern-Interaktion als gegenseitige Wechselwirkung. Somit haben theoretisch auch die (Kompetenz-)Erwartungen der Lernenden an Lehrpersonen einen Einfluss auf das Lernen und Lehren. Befunde zu Lehrenden(kompetenz)erwartungen und deren Quellen sind, wie bereits ausgeführt, zahlreich. Doch wie werden Lehrpersonen wahrgenommen? Im Rahmen des *stereotype content model* (s. Kapitel 5) werden Lehrpersonen zumeist mittelhoch bis hoch warm beschrieben (Carlsson, 2013; Cuddy et al., 2009; Eckes, 2002; Lehramtsstudierende s. Ihme & Möller, 2015). Teilweise werden Lehrpersonen als wenig(er) kompetent bezeichnet (Carlsson, 2013; Imhoff et al., 2013; Lehramtsstudierende s. Ihme & Möller, 2015), teilweise als kompetent (Cuddy et al., 2009; Eckes, 2002). Professor*innen werden in der Regel sowohl hoch kompetent als auch hoch warm beurteilt (Eckes, 2002). Somit gibt es Situationen und Lehrkontexte, in denen eine Lehrperson als Golem (geringe Kompetenzzuschreibung) oder Galatea (hohe Kompetenzzuschreibung) beschrieben werden kann.

Lernendenerwartungen an Lehrpersonen können durch unterschiedliche Mechanismen aufgebaut werden. Feldman und Prohaska (1979) listen auf, dass Erwartungen durch physische Attraktivität, Geschlecht, Ethnie und Verhaltenseigenarten der Lehrpersonen entstehen sowie durch Informationen von Geschwistern oder Freund*innen oder durch Gerüchte. Eine Reihe von Faktoren aufseiten der Lernenden, der Lehrpersonen und der Situation werden diskutiert (vgl. Casero, 2010). Wie auch bei den Lehrendenerwartungen (s. Kapitel 9) können Stereotype gegenüber Lehrpersonen zur Ausbildung von Lernendenerwartungen wirksam werden.

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_10

10.1 Geschlecht

Hinsichtlich des Geschlechts einer Lehrperson ist die Befundlage recht widersprüchlich und inkonsistent, was auch auf unterschiedliche Messinstrumente zur Bewertung und Evaluation der (vermeintlichen) Kompetenz zurückzuführen ist. Bei Joye und Wilson (2015) spiegelt sich bei der Bewertung von Professor*innen das klassische Geschlechtsstereotyp wider. So wurden Männer als bessere Erklärer (also kompetenter) und Frauen besser im Beziehungsaufbau (also wärmer) beschrieben. Dabei spielte das von der lernenden Person angenommene und nicht das tatsächliche Geschlecht eine Rolle. In einer Studie von MacNell et al. (2015) wurden das tatsächliche Geschlecht und das erwartete Geschlecht einer Lehrperson in einem Online-Kurs systematisch variiert. Dabei zeigte sich, dass das tatsächliche Geschlecht der Lehrperson keinen Einfluss auf die studentische Bewertung des Unterrichts und der Lehrperson hat. Das erwartete, angenommene Geschlecht der Lehrperson hatte hingegen einen Einfluss auf die Bewertung: So wurden vermeintlich männliche Lehrpersonen positiver bewertet als weibliche. Ein Einfluss des Geschlechts der Studierenden zur Überprüfung der Eigengruppenbevorzugung (Balliet et al., 2014; DiDonata et al., 2011; Robbins & Krueger, 2005) wurde hingegen nicht kontrolliert.

Insgesamt zeigt sich in einigen Studien, dass weibliche Lehrpersonen kompetenter und positiver bewertet werden (Bennet, 1982; Griffin, 2001; Martin & Smith, 1990; Romano & Bordieri, 1989; Sohr-Preston et al., 2016; Wigington et al., 1989), andere Studien zeigen, dass Männer besser beurteilt werden (insbesondere, wenn gleichzeitig als sympathisch resp. warm beurteilt; Basow & Silberg, 1987; Meltzer & McNulty, 2011; Nadler et al., 2013) und eine dritte Gruppe von Studien konnte keine signifikanten (Haupt-)Effekte für das Geschlecht ausmachen (Basow et al., 2013; Bavishi et al., 2010; Goebel & Cashen, 1979; Riniolo et al., 2006). In einer Untersuchung von Basow et al. (2006) nominierten Studierende zu 71 % einen männlichen Inhaber einer Professur als beste Lehrperson und zu 66 % eine weibliche Inhaberin einer Professur als schlechteste Lehrperson. Die meisten diesbezüglichen Studien fokussieren dabei auf die Hochschullehre bzw. die Beurteilung von Hochschullehrenden, was sich nicht ohne Weiteres auf andere Bildungskontexte übertragen lässt. So bewerteten Kinder bei Mancus (1992) weibliche Lehrkräfte kompetenter als männliche.

In der Zusammenschau aller rezipierten Studien werden weibliche Lehrpersonen tendenziell besser und kompetenter beurteilt als männliche. Die bessere Bewertung kann nach Griffin (2001) darauf zurückgeführt werden, dass zumeist Frauen für die Bewertung befragt werden und sich somit eine Eigengruppenbevorzugung (Balliet et al., 2014; DiDonata et al., 2011; Robbins & Krueger, 2005)

zeigt, die jedoch nicht immer ausgemacht werden kann (u. a. Basow et al., 2013). Außerdem bewerten weibliche Dozierende tendenziell milder (u. a. Nadler et al., 2013), was wiederum zu einer besseren Evaluation führen kann.

10.2 Ethnie und Migrationshintergrund

Wie bei den Erwartungen an Lernende kann auch der ethnische Hintergrund einer Lehrperson Lernendenerwartungen an die Lehrperson induzieren. Die empirischen Befunde sind auch nicht eindeutig und im Vergleich zu geschlechtsbezogenen Unterschieden eher rar.

Hinsichtlich der Kompetenzbewertung befragten Bavishi et al. (2010) High School-Schüler*innen, wie sie Professor*innen bewerten. Dafür sollten sie sich vorstellen, während des Studiums als Assistent*in für eine*einen Professor*in zu arbeiten. Die Bewertung basierte auf einem fiktiven Lebenslauf, in dem u. a. die Ethnie (afrikanisch, asiatisch oder kaukasisch) systematisch variiert wurde. Die vermeintlich afroamerikanischen Professor*innen wurden weniger kompetent bewertet als die asiatisch-amerikanischen (*Asian American*) und kaukasischen, die Bewertungen der Letzteren unterschieden sich nicht. Bei Basow et al. (2013) erhielten hingegen afroamerikanische Professor*innen dezent bessere Bewertungen als kaukasische. Glascock und Ruggiero (2006) zeigten wie Bavishi et al. (2010) für Afroamerikaner*innen, dass hispanoamerikanische Professor*innen negativer bewertet werden als angloamerikanische. Im Vergleich bezüglich der Kompetenz (*capability*) mit lateinamerikanischen Lehrpersonen schnitten angloamerikanische bei Anderson und Smith (2005) hingegen nicht vorteilhafter ab; es zeigte sich kein Haupteffekt. Insgesamt scheint also zumeist die Ethnie der Majorität (in den hier zitierten US-amerikanischen Studien die kaukasischen, angloamerikanischen Lehrpersonen) positiver bewertet zu werden.

10.3 Physische Attraktivität

Attraktive (sexuell anziehende, *hot*) Lehrpersonen werden fast durchgehend positiver bewertet und evaluiert (Bonds-Raacke & Raacke, 2007; Goebel & Cashen, 1979; Liu et al., 2013; Riniolo et al., 2006). Liu et al. (2013) beschrieben, dass der Effekt für weibliche Probandinnen größer war als für männliche Probanden. Lediglich bei Sohr-Preston et al. (2016) erhielten attraktive Dozierende geringere Werte bezüglich der Bewertung der Klarheit und ansonsten gleiche Beurteilungen. Romano und Bordieri (1989, S. 1099) resümieren:

> Analyses showed that attractive professors were (a) seen as better teachers, (b) perceived as more likely to be asked for additional assistance, (c) more likely to be recommended to other students, and (d) less likely to be blamed if a student received a failing course grade than unattractive professors.

10.4 Gerüchte und Reputation

Neben den stereotypbasierten Quellen für (Kompetenz-)Erwartungen sind auch Gerüchte und die Reputation bedeutsam. So können im schulischen Kontext Informationen zu Lehrkräften von Freund*innen und/oder den Geschwistern kommen. Stärker von Bedeutung für den Hochschulbereich sind Online-Bewertungsportale.

Einige experimentelle Studien manipulierten die Kompetenzerwartungen an Lehrpersonen via vorheriger Kommentare nach dem Muster von RateMyProfessors.com. Anschließend sollten die Proband*innen eine videografierte Lehrperson beurteilen. Die Kompetenzmanipulation beeinflusste die Beurteilung der Lehrpersonen und des Unterrichts in die zur Manipulation kongruente Richtung (Lewandowski et al., 2012; Reber et al., 2017).

Bei Wigington et al. (1989) zeigten sich u. a. die Reputation resp. das Ansehen und die berufliche Position (Rang) der dozierenden Person als Prädiktoren für die Bewertung der Lehrperson in Lehrevaluationen. Während eine hohe Reputation mit besseren Evaluationen einherging, wurden Lehrassistent*innen sowie Professor*innen schlechter bewertet als Assistenz- und assoziierte Professor*innen. Der Einfluss der Reputation zeigte sich auch bei Tollefson und Wigington (1986) sowie bei Griffin (2001).



11 Wirkungen von Lernendenkompetenzerwartungen an Lehrpersonen

Wie aufgezeigt, speisen sich die Lernendenkompetenzerwartungen (kognitives Schema nach Nickel, 1985) an Lehrpersonen aus unterschiedlichen Quellen (des soziokulturellen Bezugsrahmens, Nickel, 1985). Diese Erwartungen wirken sich nach dem transaktionalen Modell von Nickel (1985; s. Kapitel 7) sowohl auf das Erleben und Verhalten der Lernenden selbst aus als auch auf die Lehrpersonen (s. Abbildung 7.1).

11.1 Bewertung der Lehrpersonen

Ein Bereich, der augenscheinlich im Zusammenhang mit den Lernendenkompetenzerwartungen an Lehrpersonen stehen sollte, ist die Evaluation und die Bewertung der Lehrpersonen. Dies wird auch durch die empirische Befundlage gestützt.

So zeigten Terry und McIntosh (1988), dass der erste Eindruck eines Kurses inklusive des ersten Eindrucks von der Lehrperson mit der Evaluation sechs und zwölf Wochen später zusammenhängen. Für den angenommenen Lernerfolg (*amount learned*) zeigten sich keine Zusammenhänge zwischen dem ersten Eindruck und der späteren Evaluation, jedoch zwischen Evaluationen in der sechsten und zwölften Woche. Interessant ist, dass sich für die Lehr-Lern-Interaktion (*teacher's concern, methods of criticizing and correcting students*, *extent to which teacher motivates students*) zu allen drei Messzeitpunkten signifikante Zusammenhänge fanden.

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_11

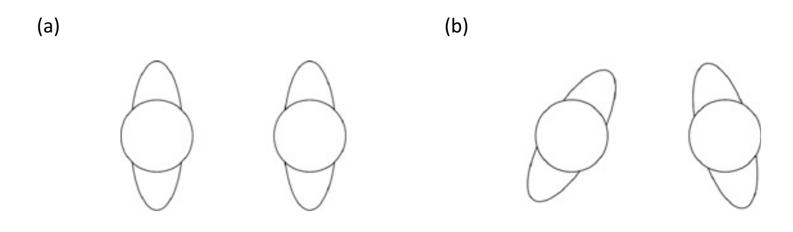
In experimentellen Studien, in denen die Lernendenkompetenzerwartungen durch Fallvignetten variiert wurden, zeigte sich, dass die spätere Bewertung der Lehrperson kompetenter, sympathischer und motivierter ausfällt, wenn die Proband*innen glaubten, sie hätten es mit einer vermeintlich kompetenten Lehrperson zu tun (Feldman & Prohaska, 1979; Feldman & Thiess, 1982; Lewandowski et al., 2012; Perry et al., 1979; Reavis, 1979; Reber et al., 2017; Towler & Dipboye, 2006). Die Effekte variieren allerdings je nach betrachteter Dimension. Haefele (1988) ließ Studierende videografierte Lehrkräfte (in einer 5. oder einer 11. Klasse) bewerten, von denen entweder behauptet wurde, sie wären kompetent (Galatea-Bedingung) oder nicht kompetent (Golem-Bedingung). Im Fokus stand dabei die Bewertung der Klarheit und Verständlichkeit. Haefele (1988) konnte jedoch keine signifikanten Unterschiede zwischen der Golem- und der Galatea-Bedingung berichten, wenngleich deskriptiv die Unterschiede in die erwartete Richtung gingen. Auch Towler und Dipboye (2006) fanden keine Unterschiede bezüglich der Klarheit, Feldman und Theiss (1982) hingegen konnten einen signifikanten Effekt ausmachen.

Leventhal et al. (1977) untersuchten Kompetenzerwartungen im Kontext der Qualität der Lehre. Sie zeigten Proband*innen entweder ein Video eines guten oder eines schlechten Vortrags. Vorher wurde die Kompetenzerwartung über die Beschreibung der Lehrperson als erfahren oder unerfahren manipuliert. Das Ergebnis zeigte einen Haupt- und einen Interaktionseffekt: Gute Vorträge erzielten insgesamt die positivsten Ergebnisse. Innerhalb der guten Vorträge wurden vermeintlich Unerfahrene schlechter in der allgemeinen Fähigkeit bewertet als Erfahrene. Kein Unterschied ergab sich zwischen den Bedingungen der schlechten Vorträge.

11.2 Erleben und Verhalten der Lernenden

Haben Lernende eine Erwartung an Lehrpersonen ausgebildet, dann beeinflusst diese auch sie selbst und dies in unterschiedlichster Weise. Feldman und Prohaska (1979, Studie 1) manipulierten die Kompetenzerwartung in einer dyadischen Lehr-Lern-Interaktion in einer Laborstudie, die auf Video aufgezeichnet wurde. In der einen Bedingung wurde behauptet, die Lehrperson (Konföderierte) sei kompetent und freundlich (Galatea-Bedingung), in der anderen, dass die Lehrperson inkompetent und uneffektiv sei (Golem-Bedingung). Dabei zeigten sich folgende Ergebnisse: Die Lernenden der Golem-Bedingung fanden (1.) den Unterricht schwieriger, (2.) den Unterricht weniger interessant und (3.) effektiv, (4.) die Lehrperson weniger kompetent, (5.) intelligent, (6.) enthusiastisch und

(7.) sympathisch als die Lernenden der Galatea-Bedingung. Weiterhin zeigten sie (8.) keine signifikant schlechtere Leistung bei Merkaufgaben (*paired associates*), (9.) die Lernenden schnitten jedoch schlechter bei einem Multiple-Choice-Test ab. (10.) Im Gesamten zeigte sich ein statistischer Trend für das nonverbale Verhalten der Lernenden: In der Galatea-Bedingung waren die Lernenden der Lehrperson zugewandter (*immediacy*). Im Einzelnen gab es einen signifikanten Unterschied bezüglich des Vorbeugens und einen Trend beim prozentualen Blickkontakt (jeweils stärkere Ausprägung in der Galatea-Bedingung). Statistisch nicht signifikant waren die Distanz zur Lehrperson und die seitliche Orientierung (Linie zwischen Schultern der Versuchsperson im schrägen Winkel (somit nicht im rechten Winkel) zur Linie zwischen den Köpfen der Versuchsperson und konföderierten Lehrperson; s. Abbildung 11.1).



Anmerkungen. Aus der Vogelperspektive stehen sich zwei Personen gegenüber, bei a ist die Orientierung zugewandter als bei b.

Abbildung 11.1 Schematische Darstellung der seitlichen Orientierung

In einer Feldstudie von Jamieson et al. (1987) wurde in zwei Klassen eine positive Erwartung hinsichtlich Kompetenz und Motivation an eine Lehrkraft induziert, zwei weitere Klassen ohne Erwartungsmanipulation dienten als Kontrollgruppe. Es zeigte sich, dass die Schüler*innen in der Experimentalgruppe nach der dreiwöchigen Unterrichtseinheit höhere akademische Leistungen erzielten als die der Kontrollgruppe. Da einige Schüler*innen sieben Tage nach der Kompetenzmanipulation auf Video aufgenommen wurden, konnte auch das nonverbale Verhalten der Kinder ausgewertet werden. Dabei zeigte sich, dass die Schüler*innen in der Experimentalgruppe mehr positives, erwünschtes (z. B. Ausführen der Arbeitsaufträge) und weniger negatives, unerwünschtes Verhalten (z. B. Unterhaltungen mit Banknachbar*innen) zeigten als die Kinder der Kontrollgruppe.

In einigen Studien zeigte sich analog zum selbstgesteuerten Lernen mit Computerprogrammen (Fries et al., 2006; vgl. Haimerl & Fries, 2010) ein Lernvorteil, wenn eine kompetente Lehrperson erwartet wird (Adediwura & Tayo, 2007; Edwards et al., 2009; Jamieson et al., 1987; Reber et al., 2017). In anderen Studien ergab sich allerdings kein Unterschied zugunsten einer hohen Lernendenkompetenzerwartung an eine Lehrperson (Lewandowski et al., 2012; Perry et al., 1979; Towler & Dipboye, 2006) und in wieder anderen nur unter bestimmten Bedingungen (z. B. bei einzelnen Testverfahren; Feldman & Prohaska, 1979; nur bei schlechten Vorträgen, Leventhal et al., 1977) oder marginal signifikant (Feldman et al., 1983). Weitergehend fanden Lernende, die eine hohe Kompetenzerwartung an eine Lehrperson hatten, den Unterricht interessanter (Feldman & Prohaska, 1979; Feldman & Theiss, 1982), hatten einen positiveren Affekt (der den Zusammenhang zwischen Kompetenzerwartung und Leistung teilweise mediiert; Edwards et al., 2009) und waren eher bereit, an einer potenziellen Lehrveranstaltung teilzunehmen (Kowai-Bell et al., 2011, 2012). Zusammenfassend stehen Lernendenkompetenzerwartungen an Lehrpersonen mit zugewandtem Verhalten und einem positiven Erleben der Lehre im Zusammenhang.

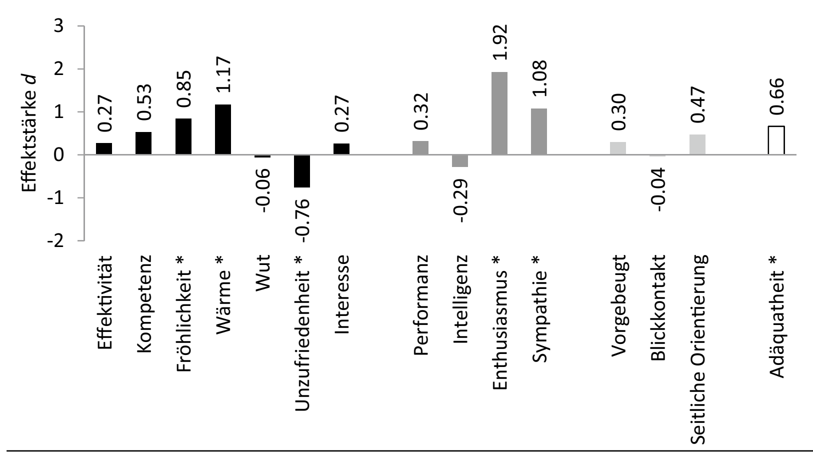
11.3 Effekte von Lernendenkompetenzerwartungen an Lehrpersonen auf die Lehrpersonen

Das von den Lehrpersonen wahrgenommene Lernendenverhalten hängt mit dem Erleben der Lehrpersonen in der Lehr-Lern-Situation zusammen. So steht die Wahrnehmung von Engagement, Disziplin und einer guten Beziehung im Zusammenhang mit der Selbstwirksamkeit, der Freude, der Angst und Frustration bei Lehrpersonen (Hagenauer et al., 2015). Wie sich die Lernendenkompetenzerwartungen auf die Lehrpersonen auswirken, ist bisher kaum gezielt empirisch untersucht, dennoch konnten einige wenige Erkenntnisse gewonnen werden. Bei Kowai-Bell et al. (2012) sollten sich Dozierende in eine auf RateMyProfessors.com bewertete Person hineinversetzen. Dabei ergab sich, dass das Hineinversetzen in eine gut bewertete Person im Vergleich zu einer schlecht bewerteten Person dazu führte, dass die Dozierenden stärker daran glaubten, dass Studierende sie mochten, dass sie mehr Selbstvertrauen im Umgang mit Studierenden in der Zukunft berichteten und einen positiveren Affekt erlebten. Kein Einfluss hingegen zeigte sich bezüglich des Vertrauens in die eigenen Fähigkeiten.

Nachdem Feldman und Prohaska (1979, Studie 1) zeigten, dass sich Lernende in Abhängigkeit ihrer Kompetenzerwartungen an die Lehrpersonen unterschiedlich verhielten, betrachteten sie die Auswirkungen der ausgemachten Verhaltensmuster auf die Lehrpersonen (Studie 2). Das nonverbale Verhalten der Lernenden

wurde von einer Konföderierten systematisch anhand der Ergebnisse der ersten Studie manipuliert. In der Galatea-Bedingung saßen die Konföderierten näher an den lehrenden Probandinnen (nur Frauen), beugten sich weiter nach vorn und hatten mehr Blickkontakt als Konföderierte der Golem-Bedingung. Außerdem saßen die Konföderierten der Galatea-Bedingung gerade ausgerichtet und die Konföderierten der Golem-Bedingung hatten eine eingedrehte Haltung (s. Abbildung 11.1). Die Effektstärken, die sich zwischen der Golem- und Galatea-Bedingung ergaben, sind in Abbildung 11.2 dargestellt. Die lehrenden Probandinnen fanden sich in der Galatea-Bedingung (1.) fröhlicher, (2.) wärmer und (3.) weniger unzufrieden mit sich selbst als die lehrenden Versuchspersonen der Golem-Bedingung. Die Lernenden wurden in der Golem-Bedingung (4.) weniger enthusiastisch und (5.) sympathisch durch die lehrenden Probandinnen eingeschätzt als in der Galatea-Bedingung. Beim nonverbalen Verhalten der Versuchspersonen gab es keine signifikanten Unterschiede. (6.) Zwei untrainierte Beobachtende, die jeweils nur 20 Sekunden einer Unterrichtseinheit auf einem Video sahen, bewerteten die lehrenden Probandinnen der Galatea-Bedingung in ihrem Verhalten signifikant adäquater als die Teilnehmenden der Golem-Bedingung. Ähnliches zeigte sich bei einer Manipulation der Erwartungen aufseiten der Lernenden hinsichtlich der Wärme (warm oder kalt; Herrell, 1971). In der Warm-Bedingung wurde die Lehrperson am Ende eines Vortrags allgemein positiver von unabhängigen Beurteiler*innen bewertet als in der Kalt-Bedingung, was auf einen Einfluss der Erwartungsmanipulation auf das Verhalten der Lehrpersonen schließen lässt.

Mottet et al. (2004) untersuchten auch die Wirkung des nonverbalen Verhaltens der Lernenden auf die Wahrnehmung von Lehrpersonen: Lehrpersonen als Studienteilnehmende sollten sich ein Video ansehen, das Lernende aus der Perspektive der Lehrperson zeigte. Anschließend sollten die Proband*innen angeben, wie sehr sie die Lernenden mochten und inwieweit sie gewillt waren, Wünschen der Lernenden nachzugeben (z. B. Fristverlängerungen). Die Videos unterschieden sich einerseits auf Basis der verbalen Aktivität der Lernenden (z. B. stellten sie viele Fragen) und im nonverbalen Verhalten der Lernenden. So zeigten die Lernenden entweder freundliches und zugewandtes (z. B. Lächeln, Blickkontakt, aufrechte Haltung) oder ein abgewendetes Verhalten (z. B. neutraler Gesichtsausdruck, wenig Blickkontakt). Es zeigte sich, dass nur das nonverbale Verhalten einen Einfluss auf die berichtete Sympathie (*liking*; 66 % erklärte Varianz) und Bereitschaft, Wünschen nachzugeben (31 % erklärte Varianz), hatte. Mottet et al. (2004) konnten also auch mit einem videobasierten Experiment die Bedeutsamkeit des nonverbalen Verhaltens von Lernenden in der Lehr-Lern-Interaktionsgestaltung aufzeigen.



Anmerkungen. Positive Werte kennzeichnen einen höheren Wert für die Galatea-Bedingung ($N = 24$) und negative Werte einen höheren Wert für die Golem-Bedingung ($N = 16$). Schwarze Balken: Selbstbewertung der Lehrpersonen; dunkelgraue Balken: Bewertung der Lernenden durch die Lehrpersonen; hellgraue Balken: nonverbales Verhalten der Lehrpersonen gegenüber den Lernenden; weiße Balken: Fremdbewertung der Lehrpersonen. * $p < .05$.

Abbildung 11.2 Effektstärken bei Feldman und Prohaska (1979)

Feldman und Theiss (1982) manipulierten ähnlich wie Feldman und Prohaska (1979) Erwartungen, aber parallel die Erwartungen der Lehrpersonen (an die Lernenden) und die der Lernenden (an die Lehrpersonen). Die Haupteffekte der Manipulation der Lehrendenkompetenzerwartung replizierte weitestgehend Ergebnisse zum Pygmalion-Effekt. Ein Effekt der Lernendenkompetenzerwartung auf die von drei Ratern durchgeführte Leistungsbewertung der Lehrperson wurde nicht signifikant. Außerdem äußerten die Proband*innen keine Unterschiede bezüglich der Einstellung der Lehrpersonen zum Unterricht oder zu den Lernenden: „Although the data reveal that teacher expectations about the student were communicated to the student, resulting in differential student performance, the results do not indicate that student expectations about the teacher were transmitted to the teacher" (Feldman & Theiss, 1982, S. 222–223). Weitergehend konnten Feldman und Theiss (1982) bei paralleler Erwartungsinduktion bei Lehrpersonen und Lernenden keine signifikanten Interaktionseffekte ausmachen.

Insgesamt zeigen sich also, sofern Unterschiede ausgemacht werden, ähnliche Phänomene wie bereits beim Pygmalion-Effekt (Rosenthal & Jacobson, 1966,

1971; vgl. Nickel, 1985), wenn die Erwartungen von Lernenden an Lehrpersonen variieren. Im Modell von Nickel (1985; s. Abbildung 7.1) lassen sich diese Befunde analog zu den Befunden der Lehrendenerwartungen einsortieren: Die Lernendenkompetenzerwartungen (als kognitive Schemata) entstehen auf Basis unterschiedlicher Informationen vor dem Hintergrund des soziokulturellen Bezugsrahmens, z. B. durch Stereotype, Bewertungen von anderen (u. a. Online-Portale) oder in Studien durch Versuchsleitungen. Die Kompetenzerwartungen führen dann zu einem angepassten Verhalten der lernenden Person, also bei hohen Kompetenzerwartungen zu einem zugewandteren und störungsärmeren Verhalten (Feldman & Prohaska, 1979; Jamieson et al., 1987). Dieses Verhalten wird von der Lehrperson wahrgenommen und in eigene Schemata in Bezug auf die lernende Person eingebaut (interaktionales Schema; Thies, 2010), die sich wiederum in eigenem Erleben und Verhalten widerspiegeln. Beispielsweise reagieren Lehrpersonen auf positive Lernendenkompetenzerwartungen mit mehr positiven und weniger negativen Emotionen, sie sind motivierter und werden effektiver im Verhalten eingeschätzt (Feldman & Prohaska, 1979). Somit können Lernendenerwartungen sowohl Ausgangspunkt für positive (Galatea-Effekt) als auch negative Lehr-Lern-Interaktionsmuster (Golem-Effekt) sein.

Die Ebene der Analysen ist bei den zitierten Studien zu den Lernendenerwartungen jedoch unterschiedlich: Einige fokussierten Erwartungen einzelner Lernender (u. a. Adediwura & Tayo, 2007; Feldman & Prohaska, 1979; Feldman & Theiss, 1982), darunter in medienbasierten Lernarrangements (u. a. Fries et al., 2006; Haimerl & Fries, 2010), andere Erwartungen von Lernendengruppen (Hagenauer et al., 2015; Herrell, 1971; Jamieson et al., 1987). Zumeist wird auf die unterschiedlichen Ebenen nicht eingegangen, aber bereits Leventhal et al. (1977) wiesen auf die Problematik der Analyse-Einheiten hin, also ob einzelne Lernende oder Gruppenwerte (*session means*) betrachtet werden sollten. Jamieson et al. (1987) diskutierten explizit, dass Lernendenerwartungen mehr Einfluss auf Lehrpersonen haben sollten, wenn die Erwartungen der Lernenden ähnlich sind – im Vergleich zu nicht geteilten, unterschiedlichen Lernendenerwartungen an Lehrpersonen. Statistisch wurde in keiner der Untersuchungen zwischen einem Einfluss auf Individual- und auf Gruppenebene unterschieden.

Im Modell von Nickel (1985) wird die Lehr-Lern-Interaktion zwischen einer lehrenden und einer lernenden Person beschrieben. Die Lernendengruppe, z. B. die Gesamtheit der Schüler*innen einer Klasse, wird nicht als Entität benannt, spielt aber im soziokulturellen Bezugsrahmen als gegenwärtige Beziehung hinein. Die Gruppendynamik innerhalb von Lehr-Lern-Gruppen ist hochkomplex und nicht-linear (Knierim et al., 2017; Koopmans & Stamovlasis, 2016; Perrez et al., 2006; Thies, 2017). So können sich Lernende an dem Verhalten von anderen

Lernenden orientieren und sich anpassen, sich den Gruppennormen widersetzen, das Gruppenklima auf die Kompetenz der Lehrpersonen zurückführen usw. (vgl. Jamieson et al., 1987) – auch Gruppenpolarisationsprozesse (vgl. Myers & Lamm, 1976) in Bezug auf die Lernendeneinschätzungen der Lehrkompetenz sind denkbar. Deswegen sollten bei Untersuchungen in Lehr-Lern-Settings sowohl die Individual- als auch die Gruppenebene betrachtet werden.

Teil IV

FRAGESTELLUNG

12 Fragestellung und Überblick über die geplanten Studien

Wie aufgezeigt, ist die Lehr-Lern-Interaktion und -Beziehung komplex und dynamisch (Knierim et al., 2017; Koopmans & Stamovlasis, 2016; Perrez et al., 2006; Thies, 2017; s. Kapitel 1) sowie wesentlich für erfolgreiches Lehren und Lernen. Auch Hattie (2015) beschreibt die Lehr-Lern-Beziehung als eine der bedeutsamsten Variablen für die Lernendenleistungen und -einstellungen (Rang 11). Unterschiedliche theoretische Ansätze zeigen die Komplexität der Lehr-Lern-Interaktion auf (vgl. Thies, 2017). Ein relevanter Faktor, der die Lehr-Lern-Interaktion beeinflussen kann, sind Erwartungshaltungen (s. Kapitel 3).

Eine besondere Form von Erwartungshaltungen wurde als Pygmalion-Effekt von Rosenthal und Jacobson (1966, 1971) beschrieben, was zu einer Reihe von Untersuchungen von Erwartungseffekten in Lehr-Lern-Situationen führte (s. Kapitel 6). Dabei wurden insbesondere interpersonale, fremdbezogene Kompetenzerwartungen von Lehrpersonen an Lernende betrachtet – sowohl in positiver (Galatea-Effekt) als auch in negativer Richtung (Golem-Effekt). Nach dem *stereotype content model* (u. a. Fiske et al., 1999; s. Kapitel 5) ist die Kompetenzbeurteilung neben der Wärmebeurteilung eine zentrale Dimension bei der sozialen Wahrnehmung, die die Lehr-Lern-Interaktion beeinflussen kann. In vielen Studien wurden experimentelle Designs eingesetzt, um den Einfluss der Kompetenzerwartungen auszumachen. Außerhalb des Labors, in „natürlichen" Lehr-Lern-Kontexten können Kompetenzerwartungen auf verschiedenen Informationen beruhen. So zeigen sich stereotypbasierte Verzerrungen der Lehrendenkompetenzerwartungen z. B. auf Basis von Geschlecht, ethnischem Hintergrund und sozioökonomischem Status der Lernenden (s. Kapitel 9).

Modelle der Lehr-Lern-Interaktion sehen allerdings nicht nur Lehrpersonen, sondern auch Lernende als aktive Gestaltende der Unterrichtssituation und der Beziehung zu den Lehrpersonen. Das transaktionale Modell von Nickel (1985; s. Kapitel 7) beschreibt eine zirkuläre Wechselwirkung der Lehrpersonen

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_12

und Lernenden, die durch individuelle kognitive Schemata, darunter Erwartungshaltungen, geprägt ist. Allerdings sind Lernendenkompetenzerwartungen an Lehrpersonen in der pädagogisch-psychologischen Forschung kaum adressiert und fokussieren meist die Auswirkungen auf die Lernenden selbst. Lediglich zwei (englischsprachige) Studien beschreiben die Wirkung von Lernendenkompetenzerwartungen auf das Erleben und Verhalten von Lehrpersonen (Feldman & Prohaska, 1979; Feldman & Theiss, 1982), die jedoch zu unterschiedlichen Ergebnissen gelangen.

Das Anliegen dieser Arbeit ist es daher, die Lernendenkompetenzerwartungen an Lehrpersonen empirisch in den Mittelpunkt zu stellen. Dabei soll nicht nur die Entstehung von Kompetenzerwartungen, sondern auch die Wirkung der Erwartungen auf die Lernenden selbst sowie auf die Lehrpersonen fokussiert werden. Dieses Anliegen führt auf Basis des transaktionalen Modells von Nickel (1985) zu drei übergeordneten Fragestellungen:

1. Wie stark werden Lernendenkompetenzerwartungen an Lehrpersonen durch (Geschlechts-)Stereotype bestimmt?
2. Wie wird das nonverbale Verhalten von Lernenden in Abhängigkeit von den Lernendenkompetenzerwartungen an Lehrpersonen beeinflusst?
3. Wie wirken initiale Lernendenkompetenzerwartungen an Lehrpersonen auf die Wahrnehmung der Lehr-Lern-Situation – sowohl bei Lernenden als auch bei Lehrpersonen?

Um diese Fragestellungen zu untersuchen, werden vier empirische Studien durchgeführt (s. Abbildung 12.1). Die Studien 1 und 2 fokussieren die erste Fragestellung. In diesen werden mithilfe von Fallvignetten die kognitiven Repräsentationen hinsichtlich der Kompetenz von schulischen Lehrkräften in Abhängigkeit von Geschlecht (Studie 1 und 2), Unterrichtsfach (Studie 1 und 2) und Schulform (Studie 2) fokussiert. Es wird davon ausgegangen, dass sich Geschlechtsstereotype, die durch den soziokulturellen Bezugsrahmen geprägt werden, in den Kompetenzbeurteilungen widerspiegeln: Basierend auf bisherigen Befunden zu Geschlechts-Professions-Assoziationen (Guimond & Roussel, 2001; Miller et al., 2018; Nosek et al., 2009; Plante et al., 2009; Smyth & Nosek, 2015; Su et al., 2009), sollten Frauen als Lehrkraft in für die Schule relativ „feminin“ angesehenen Bereichen des Unterrichts (Unterrichtsfach Deutsch; Schulform Grundschule) und Männer in relativ „maskulin“ angesehenen Bereichen (Unterrichtsfach Mathematik; Schulform Gymnasium) kompetenter beurteilt werden.

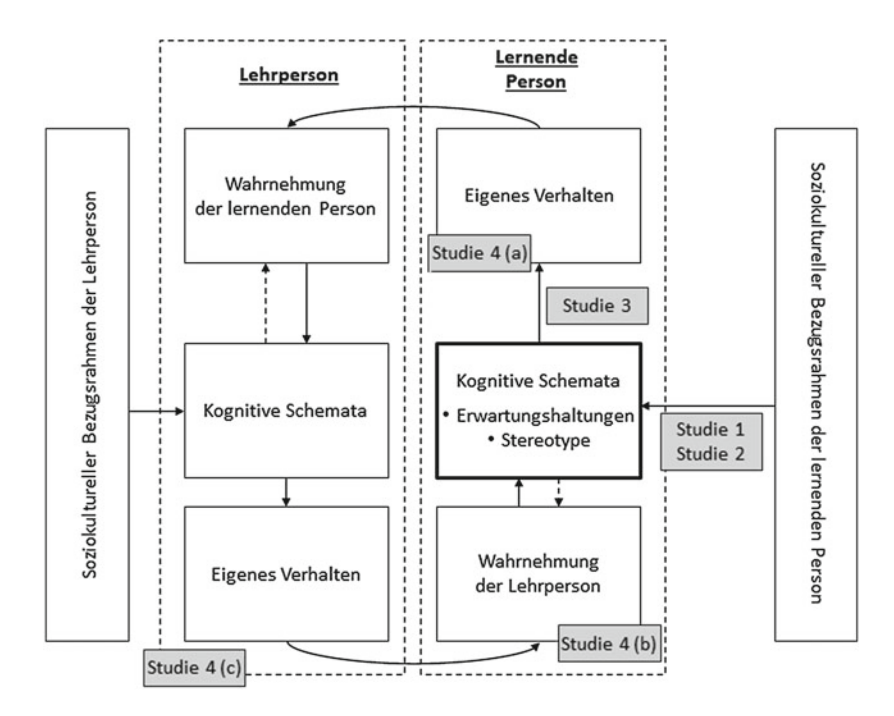


Abbildung 12.1 Überblick über die Studien in Anlehnung an das transaktionale Modell von Nickel (1985)

Studie 3 zielt auf die zweite Fragestellung. Der Pygmalion-Effekt bei Lehrendenkompetenzerwartungen wurde bereits bei Rosenthal und Jacobson (1966, 1971) dadurch erklärt, dass sich die Erwartungen im Verhalten der Lehrkräfte widerspiegeln. Diese Erklärung kann mit zahlreichen empirischen Studien (vgl. u. a. Harris & Rosenthal, 1985) Bestätigung finden. Die Befundlage zu den Auswirkungen von Lernendenkompetenzerwartungen auf das Lernendenverhalten sind rar und bereits über 30 Jahre alt (Feldman & Prohaska, 1979; Jamieson et al., 1987). In der Studie 3 werden deswegen im experimentellen Laborsetting die Kompetenzerwartungen an eine Lehrperson manipuliert. Auf Basis von videografierten Situationen werden ausgewählte Aspekte des nonverbalen Verhaltens in Abhängigkeit einer Kompetenzmanipulation überprüft.

Verschiedene weitere Zusammenhänge mit den Kompetenzerwartungen Lernender werden in der Studie 4 (zur dritten Fragestellung) in einer Feldstudie fokussiert. Dabei wird von den Lernenden der erste Eindruck, den sie von der

Lehrperson hatten, erfasst und dieser mit der Wahrnehmung des eigenen Erlebens und Verhaltens (a, Abbildung 12.1) sowie der Bewertung der Lehrperson innerhalb einer Lehrveranstaltung in Verbindung gebracht (b, Abbildung 12.1). Diese Betrachtungen sollen bisherige Befunde zur Auswirkung der Lernendenkompetenzerwartungen auf die Lernenden selbst replizieren, jedoch unter Berücksichtigung sowohl der Individual- als auch der Gruppenebene (s. Kapitel 11). Zusätzlich wird der in der bisherigen Forschung vernachlässigte Aspekt der Wirkung auf die Lehrpersonen, die bisher nur von Feldman und Prohaska (1979) sowie Feldman und Theiss (1982) untersucht wurde, empirisch überprüft (c, Abbildung 12.1). Da Feldman und Prohaska (1979) sowie Feldman und Theiss (1982) die Wirkung mit experimentellen Laborsettings sowie dyadischen Lehr-Lern-Interaktionen untersuchten, soll erweiternd und ergänzend eine Feldstudie mit Gruppen durchgeführt werden.

Teil V

EMPIRISCHE UNTERSUCHUNGEN

Check for updates

Studie 1 – Stereotypbasierte Kompetenzerwartungen an Lehrkräfte: Geschlecht und Unterrichtsfach

13

13.1 Hypothesen

Das Geschlecht einer Lehrkraft (als automatische Kategorie; Fiske & Neuberg, 1990; Krings & Kluge, 2020) kann eine Rolle spielen, um Kompetenzerwartungen bei Lernenden zu induzieren (s. Kapitel 10). Das Bild der Lehrkraft in der Schule hat sich von einem klassischen „männlichen Beruf“ zu einem „weiblichen“ entwickelt (vgl. Horstkemper, 2000) und weibliche Lehrpersonen sind zahlenmäßig in der deutlichen Mehrheit an deutschen allgemeinbildenden Schulen (73 % der Lehrkräfte sind weiblich; Statistisches Bundesamt, 2020d).

Studien auf Basis des *stereotype content model* weisen darauf hin, dass („durchschnittliche“ bzw. „typische“) Männer durchgehend wenig bis mittelhoch warm, jedoch kompetent bewertet werden (Bye et al., 2014; Durante et al., 2014; Eckes, 2002; Fiske et al., 2002). Demgegenüber werden („durchschnittliche“ bzw. „typische“) Frauen warm beurteilt, aber teils inkompetent (Eckes, 2002) oder kompetent (Bye et al., 2014; Durante et al., 2014; Fiske et al., 2002). Männern wird jedoch (deskriptiv in den erwähnten Studien) durchgehend eine höhere Kompetenz zugeschrieben als Frauen. Auch bei einer deutschen Stichprobe zeigte sich, dass Frauen warm sowie mittelhoch kompetent und Männer wenig warm und hoch kompetent beurteilt wurden (Asbrock, 2010). Unterschiedliche Fachdomänen sind mit bestimmten sozialen Gruppen verknüpft. So zeigte sich z. B., dass Assoziationen zwischen „männlich“ und Natur- und Technikwissenschaften sowie zwischen „weiblich“ und Geistes- und Sozialwissenschaften bestehen (Guimond & Roussel, 2001; Miller et al., 2018; Nosek et al. 2009;

Ergänzende Information Die elektronische Version dieses Kapitels enthält Zusatzmaterial, auf das über folgenden Link zugegriffen werden kann https://doi.org/10.1007/978-3-658-39444-8_13.

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_13

Plante et al., 2009; Smyth & Nosek, 2015; Su et al., 2009). Frauen werden mehr kommunale (z. B. kommunikativ, kooperativ, emotional) und weniger agentische Eigenschaften (z. B. autonom, dominant, selbstbewusst) zugeschrieben als Männern (u. a. Bosak et al., 2008; Diehl et al., 2004), was sich auch in den beschriebenen Ergebnissen zum *stereotype content model* zeigt und die Geschlechts-Professions-Assoziationen bedingen können.

Lehrkräfte in der Schule werden teilweise als wenig kompetent wahrgenommen (Imhoff et al., 2013; für Lehramtsstudierende s. Ihme & Möller, 2015), teilweise als kompetent (Eckes, 2002). Es stellt sich also die Frage, wie die unterschiedlichen Ergebnisse zustande kommen. Denkbar wäre, dass es unterschiedliche Gruppen von Lehrkräften gibt, die zu Substereotypisierungen (vgl. Machunsky, 2020) führen – z. B. auf Basis automatischer Kategorien (Fiske & Neuberg, 1990; Krings & Kluge, 2020). Innerhalb einer älteren Untersuchung von Mancus (1992) bewerteten Kinder weibliche Lehrkräfte kompetenter als männliche. Bei Imhoff et al. (2013) zeigten sich jedoch keine wesentlichen Unterschiede zwischen weiblichen und männlichen Lehrkräften bezüglich der Kompetenzeinschätzung. Durch die uneindeutige Befundlage ergibt sich folgende ungerichtete Hypothese, um zu überprüfen, ob Frauen oder Männer im Lehrberuf kompetenter wahrgenommen werden:

Hypothese 1: Weibliche und männliche Lehrkräfte werden (bei unbekanntem Unterrichtsfach) unterschiedlich kompetent beurteilt.

Außerdem soll parallel zum Geschlecht das Unterrichtsfach betrachtet und somit überprüft werden, ob es Geschlechts-Professions-Assoziationen (u. a. Guimond & Roussel, 2001; Miller et al., 2018; Nosek et al. 2009; Plante et al., 2009; Smyth & Nosek, 2015; Su et al., 2009) auch bei Lehrkräften gibt. Frauen als Lehrkräfte sollten demnach in mathematischen und naturwissenschaftlichen Fächern weniger kompetent beurteilt werden und Männer weniger kompetent in geistes- und sozialwissenschaftlichen Unterrichtsfächern.

Für kognitive Schemata, die sich auf schulische Lehrkräfte beziehen, zeigte sich bei Zander et al. (2015), dass Männer im Vergleich zu Frauen eher als qualifiziert genug beurteilt wurden, um Physik zu unterrichten. Weibliche Lehrkräfte, die Physik unterrichten, wurden auch optisch männlicher konstruiert als weibliche Lehrkräfte, die Deutsch unterrichten (ermittelt mit dem *reverse correlation classification task*; Degner et al., 2019). Für Physik scheint es also eine Geschlechts-Professions-Assoziation zu geben. Im Folgenden sollen als Unterrichtsfächer Mathematik und Deutsch fokussiert und verglichen werden, da beide Fächer den gleichen Stellenwert im deutschen Schulsystem einnehmen

(im quantitativen Sinn: Fach ab der ersten Klasse unterrichtet und gleiche bzw. ähnliche Anzahl Unterrichtsstunden).

Hypothese 2: Weibliche Lehrkräfte werden im Unterrichtsfach Deutsch kompetenter eingeschätzt als männliche Lehrkräfte.

Hypothese 3: Männliche Lehrkräfte werden im Unterrichtsfach Mathematik kompetenter eingeschätzt als weibliche Lehrkräfte.

Im *stereotype content model* geht in der sozialen Wahrnehmung die Wärmebeurteilung der Kompetenzbeurteilung voraus (s. Kapitel 5). Außerdem wurden bildbasierte Fallvignetten präsentiert (s. u.), sodass das *„beauty is good" stereotype* (Dion et al., 1972; Eagly et al., 1991; Langlois et al., 2000) die Bewertung der Kompetenz beeinflussen kann. Deshalb werden alle Hypothesen zusätzlich unter Kontrolle der Wärme- und Attraktivitätsbeurteilung getestet. Die Überprüfung mit Kontrollvariablen erfolgt explorativ (ungerichtet), um etwaige gegenteilige Muster und weitere Haupt- und Interaktionseffekte detektieren zu können.

13.2 Methodik

13.2.1 Stichprobe

An der Umfrage nahmen insgesamt 153 Proband*innen teil, von denen acht ausgeschlossen wurden, weil sie weniger als 50 % der Items (der nicht-soziodemografischen Variablen) beantworteten. Somit ergab sich eine Stichprobe von 145 Personen. Davon waren 99 weiblich und 40 männlich, sechs Personen machten keine Angabe zu ihrem Geschlecht. Das Durchschnittsalter betrug 28.0 Jahre ($SD = 7.38$; Spannweite: 18–52); 27 Proband*innen gaben ihr Alter nicht an. Der Großteil der Teilnehmenden studierte ($n = 91$, davon 51 Psychologie und 2 ein Lehramt), 47 Personen waren keine Studierenden (und zumeist in einem psychosozialen Beruf tätig; $n = 19$ eindeutig identifizierbar), sieben Personen machten keine Angabe.

13.2.2 Vorgehen

Es wurde eine Online-Umfrage (unipark, Questback GmbH) erstellt, die im Zeitraum vom 17.02.2015 bis zum 01.04.2015 geöffnet war. Die Proband*innen wurden über E-Mail-Verteiler für Psychologie-Studierende (in Braunschweig und

Osnabrück) sowie über soziale Medien akquiriert. Die Studie wurde als „Studie zur Einschätzung von Personen" beschrieben. Nachdem die Teilnehmenden über die Freiwilligkeit, Anonymität und den Datenschutz aufgeklärt worden waren, wurde zufällig eine von sechs Fallvignetten präsentiert, zu denen die Proband*innen ihre persönliche Einschätzung geben sollten. Die Fallvignetten bestanden aus einem Porträt, das entweder ein weibliches oder männliches Durchschnittsgesicht nach Gründl (2011[1], S. 49 weibliches Porträt, S. 21 männliches Porträt) in Graustufen und mit 25 % Rauschfilter zeigte (s. Abbildung 13.1). Bildreize wurden u. a. auch bei Martin und Smith (1990) sowie Joye und Wilson (2015) genutzt. Zusätzlich wurde mit einem Satz das dargestellte Porträt als Lehrkraft beschrieben: „Dies ist eine Lehrkraft, die an einer Supervisionsmaßnahme teilnimmt." Neben „Lehrkraft" wurden die Porträts außerdem als „Deutsch-Lehrkraft" oder „Mathematik-Lehrkraft" bezeichnet. Es handelte sich also um ein 2 × 3-Design. Anschließend wurden soziodemografische Daten erhoben und am Ende über das konkrete Anliegen der Befragung aufgeklärt sowie sich bei den Proband*innen für die Teilnahme bedankt. Die Daten wurden mittels IBM SPSS Statistics 25 und MPlus 8 ausgewertet.





Abbildung 13.1 Weibliches und männliches Durchschnittsgesicht (modifiziert nach Gründl, 2011; Studie 1)

[1] Mit freundlicher Genehmigung von Martin Gründl.

13.2.3 Instrumente und Variablen

Als abhängige Variable wurde die Beurteilung der Kompetenz erfragt und die Beurteilung der Wärme sowie der Attraktivität wurden als Kontrollvariablen herangezogen (Fragebogen s. Zusatzmaterial).

Kompetenzbeurteilung: Die Einschätzung der Kompetenz wurde mittels einer Adjektivliste erfasst, bei der die Teilnehmenden auf einer vierstufigen Likert-Skala (1 = „stimme gar nicht zu" bis 4 = „stimme vollkommen zu") angeben sollten, inwieweit das Adjektiv auf die dargestellte Person der Fallvignette zutrifft. Hierfür wurden folgende Adjektive herangezogen (angelehnt an Fiske et al., 1999; Imhof et al., 2013; Rosenberg et al., 1968): kompetent, fähig, effizient, intelligent, konkurrenzfähig, selbstbewusst und eigenständig. Die Reihenfolge der Items wurde zufällig präsentiert. Nach Faktorenanalysen und aus inhaltlichen Gründen (s. u.) wurden die Items konkurrenzfähig, selbstbewusst und eigenständig nicht berücksichtigt. Die Reliabilität (Cronbachs Alpha) der 4-Item-Lösung liegt mit $\alpha = .831$ in einem zufriedenstellenden Bereich.

Wärmebeurteilung: Die Einschätzung der Wärme wurde analog mit einer Adjektivliste und einer vierstufigen Likert-Skala wie die Beurteilung der Kompetenz erfasst. Die Items (angelehnt an Fiske et al., 1999; Imhof et al., 2013; Rosenberg et al., 1968) wurden zusammen in zufälliger Reihenfolge mit denen der Kompetenzbeurteilung präsentiert: warmherzig, gutmütig, wohlwollend, sympathisch, offen, tolerant und vertrauenswürdig. Die Items offen, tolerant und vertrauenswürdig wurden auf Basis von inhaltlichen und faktoranalytischen Betrachtungen (s. u.) final nicht berücksichtigt. Die interne Konsistenz (Cronbachs Alpha) der 4-Item-Lösung liegt in einem zufriedenstellenden Bereich mit $\alpha = .856$. Die Wärmebeurteilung wird im Sinne des *stereotype content model* als Kontrollvariable herangezogen.

Attraktivitätsbeurteilung: Die Bewertung der Attraktivität wurde mit folgendem Item auf einer siebenstufigen Likert-Skala (1 = „sehr unattraktiv" bis 7 = „sehr attraktiv") erhoben: „Wie optisch attraktiv finden Sie die Lehrkraft?" Die Erfassung der Attraktivitätseinschätzung mit einem Item ist nicht unüblich (z. B. bei Kemper et al., 2014). Die Attraktivitätsbeurteilung wird berücksichtigt, um das *„beauty is good" stereotype* aufgrund des Bildreizes zu kontrollieren.

Eine konfirmatorische Faktorenanalyse mit robuster Maximum-Likelihood-Schätzung (MLR) zeigte für eine 2-Faktoren-Lösung (Kompetenz- und Wärmebeurteilung) mit jeweils sieben Items der Kompetenz- und Wärmebeurteilung ungenügende Fitmaße ($\chi^2/df = 2.545$; CFI = .857; TLI = .829; RMSEA = .103; SRMR = .092; Hu & Bentler, 1999; Weiber & Mühlhaus, 2014). Ein einfaktorielles Modell ergab eine schlechtere Lösung ($\chi^2/df = 4.996$;

CFI = .626; TLI = .558; RMSEA = .166; SRMR = .134). Aus inhaltlichen Gründen wurden anschließend die Items „vertrauenswürdig“, „konkurrenzfähig“ und „eigenständig“ aus den Analysen herausgenommen. Die Vertrauenswürdigkeit kann u. U. bei Lehrpersonen auch auf fachliche Kompetenz hinweisen. Konkurrenzfähigkeit und Eigenständigkeit könnten ungeeignet sein, um die Kompetenz von Lehrpersonen zu erfassen. Danach wurden zusätzlich schrittweise Items auf Basis der Trennschärfe entfernt, sodass jeweils vier Items pro Faktor verblieben[2] (Itemschwierigkeiten und Trennschärfen s. Zusatzmaterial). Die endgültige, ökonomische zweifaktorielle Lösung ergab zufriedenstellende Fitmaße ($\chi^2/df = 1.423$; CFI = .980; TLI = .971; RMSEA = .054; SRMR = .049). Ein alternativ getestetes einfaktorielles Modell mit allen acht verbliebenen Items war zu verwerfen ($\chi^2/df = 8.889$; CFI = .623; TLI = .472; RMSEA = .233; SRMR = .122). Die 2-Faktoren-Lösung mit jeweils vier Items konnte auch an anderer Stelle zur Beurteilung von Erziehungsberatenden zufriedenstellende Fitmaße erzielen (Hackbart et al., 2020). Aufgrund der geringen Abstufungen der Likert-Skala (Bühner, 2010) wurde die finale Faktorenstruktur zusätzlich mit einem auch für ordinale Daten angemessenen Weighted-Least-Squares-Verfahren (WLSMV; Eid et al., 2017; Holtmann et al., 2016; Li, 2016) mit Ausnahme des RMSEA bestätigt ($\chi^2/df = 2.547$; CFI = .976; TLI = .964; RMSEA = .103; WRMR[3] = 0.712). Der RMSEA fällt jedoch bei kleinen Stichproben ($N < 250$) häufig hoch aus (vgl. Weiber & Mühlhaus, 2014).

13.3 Ergebnisse

Die Kennwerte der Variablen sowie die Korrelationen untereinander sind in Tabelle 13.1 aufgelistet, die Kennwerte nach Gruppen in Tabelle 13.2. Da einige der Variablen hinsichtlich der Normalverteilung auffällig sind (auf Basis von Shapiro-Wilk-Tests), wurden sowohl parametrische als auch nicht-parametrische Korrelationen berechnet. Die Quantil-Quantil-Diagramme sind bei optischer Inspektion jedoch unauffällig.

[2] Die Reduzierung der Items wurde teilweise trotz guter Trennschärfen vorgenommen, um für die folgenden Studien eine ökonomische Adjektivliste zu erhalten.

[3] Weighted Root Mean Square Residual

Tabelle 13.1 Kennwerte und Korrelationen der Variablen (Studie 1)

	N	*M*	*SD*	α	SW	Korrelationen[a] 1	2	3	4	5	6	7	8	9
*Proband*innen*														
1 Geschlecht[b]	139	1.71	0.45		.57**	–	−.21	.04	−.01	−.03	.01	.01	−.14	.14
2 Alter	118	28.04	7.38		.88**	−.09	–	−.10	−.14	.19*	.08	.05	.10	−.17
Beurteilungen														
3 Kompetenzbeurteilung[c]	143	2.82	0.58	.83	.95**	.01	−.09		.34**	.26**	.31**	−.18*	.05	.13
4 Wärmebeurteilung[c]	142	2.62	0.66	.86	.97**	−.01	−.18	.43**	–	.29**	.04	−.01	−.02	.04
5 Attraktivitätsbeurteilung[d]	144	4.73	1.12		.92**	−.05	18*	.27**	.31**	–	.22**	.05	.21*	.16*
Bedingungen														
6 Geschlecht der LK[b]	145	1.54	0.50		.63**	.01	.11	.27**	.05	.23**	–	−.08	.01	.07
7 Fach Deutsch[e]	145	0.33	0.47		.59**	.01	.08	−.15	−.01	.01	−.08	–	.53**	.46**
8 Fach Mathematik[e]	145	0.37	0.48		.61**	−.14	.06	.04	−.05	−.19*	.01	.53**	–	.50**
9 Unbekanntes Fach[e]	145	0.30	0.46		.58**	.14	−.15	.11	.06	.19*	.07	.46**	.50**	–

Anmerkungen. α = Cronbachs Alpha; SW = Shapiro-Wilk-Test auf Normalverteilung; LK = Lehrkraft; Fach = Unterrichtsfach. [a] Korrelationen: $115 \leq N \leq 145$; über der Diagonalen: Korrelationen nach Spearman; unter der Diagonalen: Korrelationen nach Pearson. [b] Codierung: 1 = männlich, 2 = weiblich. [c] Mögliche Werte: 1–4. [c] Mögliche Werte: 1–7. [e] Dichotomisierte Gruppenzuordnung für das Unterrichtsfach der Lehrkraft, Codierung: 0 = nein, 1 = ja. *$p < .05$. ** $p < .01$.

Die bivariaten Analysen zur Überprüfung der Hypothesen wurden mit t-Tests durchgeführt. Weibliche Lehrkräfte ohne bekanntes Unterrichtsfach (M = 3.106, SD = 0.575) wurden signifikant kompetenter beurteilt als männliche Lehrkräfte in einem unbekannten Unterrichtsfach (M = 2.639, SD = 0.544; $t(42)$ = 2.706, p = .010; d = 0.830; 1–β = .753)[4]. Auch über alle Bedingungen der Unterrichtsfächer hinweg (Deutsch, Mathematik und unbekanntes Unterrichtsfach) waren die Kompetenzbeurteilungen für weibliche Lehrkräfte höher (M = 2.964, SD = 0.586) als für männliche (M = 2.652, SD = 0.539; $t(141)$ = 3.302, p = .001; d = 0.554; 1–β = .906). Für den Fachunterricht ergab sich, dass weibliche Deutschlehrkräfte signifikant kompetenter beurteilt wurden (M = 2.891, SD = 0.532) als männliche (M = 2.520, SD = 0.478; $t(46)$ = 2.546, p_1 = .007; d = 0.736; 1–β = .806). Kein Geschlechtsunterschied ergab sich für den Mathematikunterricht ($t(49)$ = 0.516, p_1 = .304; d = 0.145; 1–β = .128).

Für die multivariate Berechnung der Gruppenunterschiede unter Berücksichtigung des *stereotype content model* und des *„beauty is good" stereotype* wurde eine Kovarianzanalyse durchgeführt, bei der die Kompetenzbeurteilung als abhängige Variable einging und das Geschlecht sowie das Unterrichtsfach der Lehrkraft als unabhängige Variable. Die Attraktivitäts- und Wärmebeurteilung gingen als Kontrollvariablen ein. Die Kovarianzanalyse ist ein robustes Verfahren, auch gegenüber Voraussetzungsverletzungen (Harwell, 2003; vgl. Bortz & Schuster, 2010). Nichtsdestotrotz soll kurz auf die Eignung des vorliegenden Datensatzes eingegangen werden. Die Prämissen für die hypothesenprüfende Kovarianzanalyse wurden in Anlehnung an Huber et al. (2014) überprüft: (1.) Keine Ausreißer: Auf einen Ausschluss von Ausreißern wurde verzichtet, da ausschließlich geschlossene Skalen vorlagen. (2.) Eine randomisierte Gruppenzuordnung lag vor. (3.) Gruppengröße über 20: Eine Gruppe (männliche Lehrkraft ohne genanntes Fach) liegt mit 18 Proband*innen leicht unter den empfohlenen 20 Teilnehmenden pro Gruppe. (4.) Die Varianzhomogenität war gemäß dem Levene-Test gegeben ($F(5, 133)$ = 0.545, p = .742). (5.) Die Normalverteilung der Variablen ist teils gegeben, teils nicht (abweichend von Huber et al., 2014, mit dem Shapiro-Wilk-Test berechnet; s. Tabelle 13.2). (6.) Die Kovariaten bzw. die Kontrollvariablen korrelieren mit der abhängigen Variablen. (7.) Es gab keine signifikanten Interaktionseffekte einer unabhängigen Variablen mit einer

[4] Das Signifikanzniveau für gerichtete Hypothesen bzw. einseitige Tests wird mit p_1 angegeben, für ungerichtete Hypothesen bzw. für zweiseitige Tests mit p. Ein Signifikanzniveau, das sich aus der Überprüfung mit Bootstrap-Verfahren ergibt, wird mit p_{bt} (zweiseitig) angegeben. Alle angegebenen Powerwerte (1–β) wurden für die erzielten Effektstärken über Post-hoc-Analysen (α = .05) mit der Software G*Power 3.1 bestimmt.

Tabelle 13.2 Mittelwerte, Standardabweichungen und Normalverteilungsüberprüfungen der Skalen getrennt nach Gruppen (Studie 1)

		Kompetenzbeurteilung[a]			Wärmebeurteilung[a]			Attraktivitätsbeurteilung[b]		
Gruppe	*N*	*M*	*SD*	SW	*M*	*SD*	SW	*M*	*SD*	SW
Alle Gruppen	145	2.82	0.58	.95**	2.62	0.66	.97**	4.73	1.12	.92**
Männliche LK	67	2.65	0.54	.96*	2.58	0.59	.97	4.45	1.06	.92**
Weibliche LK	78	2.96	0.59	.91**	2.65	0.73	.97*	4.97	1.12	.91**
Deutsch-LK	48	2.70	0.53	.95*	2.61	0.64	.94*	4.75	1.23	.92**
Mathematik-LK	53	2.85	0.60	.95*	2.58	0.69	.97	4.45	1.01	.90**
LK ohne Fach	44	2.91	0.60	.93**	2.67	0.66	.96	5.05	1.07	.88**
Männliche Deutsch-LK	25	2.52	0.48	.95	2.53	0.64	.94	4.44	1.29	.89*
Männliche Mathematik-LK	24	2.80	0.58	.95	2.58	0.50	.93	4.21	0.78	.86**
Männliche LK ohne Fach	18	2.64	0.54	.90	2.67	0.64	.91*	4.78	1.00	.90
Weibliche Deutsch-LK	23	2.89	0.53	.91*	2.70	0.65	.89*	5.09	1.08	.91*
Weibliche Mathematik-LK	29	2.89	0.63	.94	2.58	0.82	.96	4.66	1.14	.92*
Weibliche LK ohne Fach	26	3.11	0.58	.83**	2.68	0.69	.94	5.24	1.09	.86**

Anmerkungen. N: teils geringfügige Abweichung aufgrund fehlender Werte möglich. SW = Shapiro-Wilk-Test auf Normalverteilung; LK = Lehrkraft; Fach = Unterrichtsfach. [a] Mögliche Werte: 1–4. [b] Mögliche Werte: 1–7. * $p < .05$. ** $p < .01$.

Kontrollvariablen auf die Kompetenzbeurteilung. (8.) Homogenität der Kontrollvariablen über die Experimentalgruppen: Da die Relation von kleinster zu größter Experimentalgruppe mit 1.56 (18 zu 28) nur leicht über dem empfohlenen Größenunterschied von 1.5 liegt, wurde auf die Überprüfung des Einflusses der Kontrollvariablen auf die Experimentalgruppen verzichtet, da von einer relativen Gleichbesetzung der Zellen ausgegangen werden kann. Aufgrund der geringen Abweichungen und unter Anbetracht der Robustheit von varianzanalytischen Verfahren wurde die geplante Kovarianzanalyse durchgeführt.

Zusätzlich zur Überprüfung der Voraussetzungen der Kovarianzanalyse wurde eine potenzielle Eigengruppenbevorzugung (vgl. Balliet et al., 2014; DiDonato et al., 2011; Robbins & Krueger, 2005) auf Basis des Geschlechts in einer separaten Varianzanalyse betrachtet. Es sollte betrachtet werden, ob Männer männliche Lehrkräfte besser bewerteten und Frauen weibliche Lehrkräfte. Die Varianzanalyse mit dem Geschlecht der beschriebenen Lehrkräfte und der Proband*innen als unabhängige Variablen ergab einen signifikanten Effekt des Geschlechts der Lehrkraft ($\eta_p^2 = .067$, $p = .002$)[5], aber weder einen signifikanten Haupteffekt des Geschlechts der Proband*innen ($\eta_p^2 = .000$, $p = .917$) noch eine signifikante Interaktion ($\eta_p^2 = .005$, $p = .428$; s. Zusatzmaterial). Somit kann eine Eigengruppenbevorzugung auf Basis des Geschlechts ausgeschlossen werden.

In Tabelle 13.3 sind die Ergebnisse der hypothesenprüfenden Varianzanalyse mit Kontrollvariablen dargestellt. Es zeigte sich in der Kovarianzanalyse, dass die Wärmebeurteilung ein signifikanter Prädiktor für die Kompetenzbeurteilung ist ($F(1, 131) = 24.722$, $p < .001$; $\eta_p^2 = .159$), die Attraktivitätsbeurteilung hingegen nicht ($F(1, 131) = 1.260$, $p = .264$; $\eta_p^2 = .010$). Weitergehend ergaben sich Unterschiede hinsichtlich des Geschlechts der Lehrkraft ($F(1, 131) = 7.896$, $p = .006$). Die Effektstärke ($\eta_p^2 = .057$) weist auf einen kleinen Effekt hin. Bei Inspektion der Konfidenzintervalle zeigte sich (unter Kontrolle der Wärme- und Attraktivitätsbeurteilung; s. Tabelle 13.4), dass weibliche Lehrkräfte ($M = 2.94$; 95%–CI = 2.82–3.06) kompetenter eingeschätzt wurden als männliche ($M = 2.69$; 95%–CI = 2.56–2.80).

[5] Das Effektstärkemaß partielles Eta-Quadrat wird mit η_p^2 angegeben.

Tabelle 13.3 Ergebnisse der Kovarianzanalyse zur Aufklärung der Kompetenzbeurteilung (Studie 1)

Quelle der Varianz	*df*	*F*	*p*	Partielles η^2
Korrigiertes Modell	7	7.456	< .001	.285
Konstanter Term	1	54.767	< .001	.295
Kontrollvariablen				
Wärmebeurteilung	1	24.722	< .001	.159
Attraktivitätsbeurteilung	1	1.260	.264	.010
Bedingungen				
Geschlecht der LK	1	7.896	.006	.057
Unterrichtsfach der LK	2	1.536	.219	.023
Interaktion				
Geschlecht der LK x Unterrichtsfach der LK	2	1.774	.174	.026
Fehler	131			

Anmerkungen. Korrigiertes $R^2 = .247$ ($R^2 = .285$). $N = 139$. LK = Lehrkraft.

Das Unterrichtsfach der Lehrkraft hatte keinen signifikanten Einfluss auf die Kompetenzbeurteilung ($F(2, 131) = 1.536$, $p = .219$; $\eta_p^2 = .023$). Außerdem gab es keine signifikante Interaktion von Geschlecht und Unterrichtsfach der Lehrkraft ($F(2, 131) = 1.774$, $p = .174$; $\eta_p^2 = .026$). Beim deskriptiven Vergleich der sechs Gruppen anhand der Konfidenzintervalle (bei Kontrolle für Wärme- und Attraktivitätsbeurteilung; s. Tabelle 13.4), zeigte sich, dass weibliche Lehrkräfte mit unbekanntem Unterrichtsfach ($M = 3.04$; 95%–CI = 2.83–3.25) kompetenter eingeschätzt wurden als männliche Deutschlehrkräfte ($M = 2.57$; 95%–CI = 2.36–2.77).

Tabelle 13.4 Geschätzte Randmittel unter Berücksichtigung von Kontrollvariablen (Studie 1)

					95%–CI	
Gruppenzuordnung		*N*	*M*	*SE*	*LL*	*UL*
Geschlecht der LK						
Männlich		65	2.690	0.064	2.563	2.818
Weiblich		74	2.942	0.060	2.823	3.061
Unterrichtsfach der LK						
Deutsch		47	2.719	0.074	2.572	2.865
Mathematik		50	2.900	0.073	2.755	3.045
Ohne Fach		42	2.829	0.080	2.671	2.987
Geschlecht der LK	Unterrichtsfach der LK					
Männlich	Deutsch	25	2.565	0.102	2.363	2.767
	Mathematik	22	2.885	0.110	2.667	3.102
	Ohne Fach	18	2.621	0.120	2.385	2.858
Weiblich	Deutsch	22	2.873	0.109	2.657	3.089
	Mathematik	28	2.915	0.096	2.726	3.105
	Ohne Fach	24	3.037	0.106	2.828	3.246

Anmerkungen. Abhängige Variable: Kompetenzbeurteilung. Die Kontrollvariablen im Modell wurden anhand der folgenden Werte berechnet: Wärmebeurteilung = 2.617, Attraktivitätsbeurteilung = 4.763. Anpassung für Mehrfachvergleiche nach Bonferroni. LK = Lehrkraft.

13.4 Zusammenfassung, Interpretationen und Limitationen

In der durchgeführten Online-Untersuchung sollte herausgefunden werden, ob es Substereotype gegenüber bestimmten Gruppen von Lehrkräften gibt bzw. ob sich Geschlechts-Professions-Assoziationen auch bei Lehrkräften wiederfinden. Diese Stereotype und Assoziationen sollten gemäß dem transaktionalen Modell von Nickel (1985) durch den soziokulturellen Kontext geprägt sein sowie mit Kompetenzerwartungen einhergehen.

Es zeigte sich auf Basis einer durchgeführten Studie mit Fallvignetten, dass weibliche Lehrkräfte kompetenter beurteilt werden als männliche. Dies widerspricht dem Befund von Imhoff et al. (2013), die keine Unterschiede bei der Beurteilung von männlichen und weiblichen Lehrkräften fanden, und unterstützt die Ergebnisse von Mancus (1992). Eine methodische Erklärung zur positiveren Kompetenzbewertung der weiblichen Lehrkräfte im Vergleich zu den männlichen kann im Stimulus-Material und dem *„beauty is good“ stereotype* (Dion et al., 1972; Eagly et al., 1991; Langlois et al., 2000) liegen. So gingen für das Bild der weiblichen Lehrkraft 64 Bilder zum Erstellen des Porträts und für die männliche Lehrkraft 32 Bilder ein, was zu einer höheren Durchschnittlichkeit und somit zu einer höheren Attraktivitätsbeurteilung führen kann bzw. es gingen mehr attraktive Bilder in das weibliche Gesicht ein als in das männliche, was die Attraktivitätszuschreibung auch erhöhen kann (vgl. Gründl, 2011). Zudem wurden bei Joye und Wilson (2015) Bilder von Frauen attraktiver bewertet als Bilder von Männern. Es zeigte sich ein kleiner korrelativer Zusammenhang zwischen Geschlecht der Lehrkraft und Attraktivitätsbeurteilung zugunsten der weiblichen Lehrkräfte sowie darüber hinaus ein weiterer kleiner Zusammenhang zwischen der Kompetenzbeurteilung und der Attraktivitätsbeurteilung, was auf einen mediierenden Effekt hinweisen kann.

Bei einer gemeinsamen Berücksichtigung von Geschlecht der Lehrkraft, Unterrichtsfach, Attraktivitäts- und Wärmebeurteilung verliert sich der signifikante Effekt der Attraktivitätsbeurteilung. Lediglich das Geschlecht der Lehrkraft und die Wärmebeurteilung können die Kompetenzbeurteilung signifikant vorhersagen. Besondere Bedeutsamkeit für die Kompetenzbeurteilung mit einem mittelgroßen Effekt hat die Wärmebeurteilung ($\eta_p^2 = .159$). Dies kann damit interpretiert werden, dass das stereotype Bild von guten, kompetenten Lehrkräften ein sympathisches und fürsorgliches Verhalten einschließt. Generell werden Lehrkräfte als besonders warm beschrieben (Eckes, 2002; Imhoff et al., 2013; für Lehramtsstudierende s. Ihme & Möller, 2015).

Es ist zu vermuten, dass (gerade bei unbekanntem Fach) der Erziehungsauftrag der Lehrkräfte stärker im Fokus steht als der Auftrag, Wissen zu vermitteln, und somit die Kompetenz von Lehrkräften über die Erziehungsfähigkeiten beurteilt wird. Die stärkere Zuweisung kommunaler Eigenschaften für Frauen im Verhältnis zu Männern (u. a. Bosak et al., 2008; Diehl et al., 2004) kann somit die höhere Kompetenzbeurteilung erklären. Allgemein ergreifen Frauen mehr Berufe im erzieherischen und pflegenden Bereich als Männer (Autorengruppe Fachkräftebarometer, 2019; Bundesagentur für Arbeit, 2021) und studieren u. a. häufiger Erziehungswissenschaften (79 %) oder mit dem Ziel eines Lehramts

(71 %; Statistisches Bundesamt, 2020c). Für die geschlechtsabhängige Kompetenzbeurteilung von Lehrkräften muss jedoch festgehalten werden, dass in der vorliegenden Untersuchung lediglich ein kleiner Effekt ($\eta_p^2 = .057$) erzielt wurde.

Die fehlende Signifikanz hinsichtlich der Interaktion von Unterrichtsfach und Geschlecht der Lehrkraft in den multivariaten Analysen widerspricht den Hypothesen 2 und 3. Allerdings zeigte sich in den bivariaten Analysen, das weibliche Deutschlehrkräfte kompetenter beurteilt werden als männliche Deutschlehrkräfte. Bisherige Untersuchungen, die den Deutsch- und Physikunterricht fokussierten, konnten auch Geschlechts-Professions-Assoziationen ausmachen (Degner et al., 2019; Zander et al., 2015). Es kann vermutet werden, dass die Fachauswahl bedeutsam ist. So kann der Mathematikunterricht in der Schule weniger männlich assoziiert sein als der Physikunterricht. Bei Plante et al. (2009) zeigte sich, dass Schüler*innen Mathematik nicht als männliche Domäne assoziieren, aber den Sprachunterricht als weibliche. Eine systematische Analyse der Geschlechts-Professions-Assoziationen für die Gruppe der Lehrkräfte (und andere Lehrpersonen) wäre künftig wünschenswert.

Auf Basis des Untersuchungsdesigns ergeben sich einige Limitationen. Bei der Erhebung wurde auf eine Gelegenheitsstichprobe zurückgegriffen, die hinsichtlich der Interpretierbarkeit der Ergebnisse für eine Zielpopulation begrenzt ist (vgl. Döring & Bortz, 2016). Dies macht sich in der hohen Anzahl von Studierenden und von Proband*innen aus dem psychosozialen Bereich bemerkbar, die nicht ohne Weiteres als repräsentativ für die Allgemeinbevölkerung angesehen werden können. Außerdem muss erwähnt werden, dass die Stichprobe aus erwachsenen Proband*innen bestand (Mindestalter = 18). Es kann jedoch davon ausgegangen werden, dass die Teilnehmenden auf Basis ihrer aufgebauten Schemata und Stereotype, die sich im Laufe der Zeit herausgebildet haben (soziale Lernvergangenheit; Nickel, 1985), die beschriebenen Lehrkräfte beurteilt haben. Inwieweit diese Ergebnisse auch für aktuell schulpflichtige Kinder und Jugendliche erzielt werden können, bleibt auf Basis der vorliegenden Daten offen und müsste künftig in der Forschung adressiert werden. Schließend muss noch erwähnt werden, dass theoriebasierte Items eigenständig konstruiert wurden. Diese und die daraus entstandenen Skalen sind deswegen nicht validiert, die Reliabilitäten und die faktorielle Validität können jedoch als angemessen angesehen werden. Problematisch an den Items kann auch sein, dass sie ausschließlich positiv gepolt waren (vgl. u. a. Bühner, 2010; Döring & Bortz, 2016).

Zusammenfassend kann festgehalten werden, dass weibliche Lehrkräfte kompetenter beurteilt werden als männliche. Dies spricht für die Entwicklung von

(relationalen) Schemata bzw. Substereotypen. Um den Einfluss der Attraktivität methodisch zu eliminieren und der Hypothese nachzugehen, dass weibliche Lehrkräfte insbesondere bei erzieherisch assoziierten Aufgaben kompetenter eingeschätzt werden als männliche Lehrkräfte, wurde eine zweite Studie mit textbasierten Fallvignetten erstellt. Mit der zweiten Studie sollten außerdem die in der ersten Studie erzielten Ergebnisse, insbesondere zu den Geschlechts-Professions-Assoziationen, erneut überprüft werden.



Studie 2 – Stereotypbasierte Kompetenzerwartungen an Lehrkräfte: Geschlecht, Unterrichtsfach und Schulform

14

14.1 Hypothesen

Die zweite Studie sollte weitestgehend die Hypothesen zu Geschlechts-Professions-Assoziationen der ersten Studie (s. Kapitel 13) mit einem bildreizunabhängigen Design erneut überprüfen.

Hypothese 1: Weibliche Lehrkräfte werden im Unterrichtsfach Deutsch kompetenter eingeschätzt als männliche Lehrkräfte.

Hypothese 2: Männliche Lehrkräfte werden im Unterrichtsfach Mathematik kompetenter eingeschätzt als weibliche Lehrkräfte.

Zusätzlich wurde die Schulform der Lehrkräfte berücksichtigt. Der Lehrberuf an allgemeinbildenden Schulen ist beruflich eine „weibliche" Domäne (vgl. Horstkemper, 2000; s. Kapitel 13). Der Anteil der Frauen im Lehrberuf variiert allerdings je nach Schul- und Unterrichtsform und spiegelt das Bild wider, dass die größten Frauenanteile eher in Berufen der Pflege und Erziehung zu finden sind (Autorengruppe Fachkräftebarometer, 2019; Bundesagentur für Arbeit, 2021), wo kommunale Eigenschaften (z. B. kommunikativ, kooperativ, emotional), die eher Frauen zugeschrieben werden (u. a. Bosak et al., 2008; Diehl et al., 2004; Eagly & Mladinic, 1994), besonders relevant sind. So sind zwar 73 % der Lehrkräfte weiblich, allerdings 89 % in Grundschulen, 81 % in schulartunabhängigen

Ergänzende Information Die elektronische Version dieses Kapitels enthält Zusatzmaterial, auf das über folgenden Link zugegriffen werden kann https://doi.org/10.1007/978-3-658-39444-8_14.

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_14

Orientierungsstufen, 78 % in Förderschulen, 76 % in Schulen des Gesundheitswesens, 66 % in Realschulen, 65 % in Hauptschulen, 61 % in Gymnasien und 53 % in beruflichen Schulen (Statistisches Bundesamt, 2020d). In Hochschulen liegt der Anteil der Frauen bei 40 % (hauptberufliches wissenschaftliches und künstlerisches Personal; Statistisches Bundesamt, 2020b).

Für den Lehr-Lern-Kontext sollten also nun stereotypkonform in Bereichen, in denen erzieherische Aspekte stärker im Fokus stehen (z. B. in der Grundschule), Frauen als geeigneter angesehen werden als Männer. Männer hingegen sollten als geeigneter oder mindestens als genauso geeignet beurteilt werden wie Frauen, wenn der Fokus weniger auf erzieherischen und stärker auf fachlichen Aspekten liegt (z. B. im Gymnasium). Dies sollte sich auch in der Kompetenzbeurteilung niederschlagen:

Hypothese 3: Männliche Lehrkräfte werden im Gymnasium kompetenter eingeschätzt als in der Grundschule.

Hypothese 4: Weibliche Lehrkräfte werden in der Grundschule kompetenter eingeschätzt als im Gymnasium.

Hypothese 5: Weibliche Lehrkräfte werden in der Grundschule kompetenter eingeschätzt als männliche Lehrkräfte in der Grundschule.

Wie in der ersten Studie soll auf Basis des *stereotype content model* (s. Kapitel 5) für die Wärmebeurteilung, die der Kompetenzbeurteilung vorausgeht, kontrolliert und die Hypothesen sowie weitere Haupt- und Interaktionseffekte explorativ gemeinsam betrachtet werden.

14.2 Methodik

14.2.1 Stichprobe

An der Umfrage nahmen insgesamt 284 Proband*innen teil, von denen 19 ausgeschlossen wurden, weil sie weniger als 50 % der Items (der nicht-soziodemografischen Variablen) beantworteten, und weitere elf Proband*innen, weil sie an der ersten Studie teilnahmen. Somit ergab sich eine Stichprobe von 254 Personen. Davon waren 206 weiblich und 46 männlich, zwei Personen machten keine Angabe zu ihrem Geschlecht. Das Durchschnittsalter war mit 24.9 Jahren ($SD = 5.88$; Spannweite: 18–60) unter dem der ersten Studie. Das Alter wurde von einer teilnehmenden Person nicht angegeben. Der Großteil der Proband*innen studierte ($n = 224$, davon 58 Psychologie und 78 ein Lehramt), 24 Personen waren keine Studierenden (psychosozialer Beruf: $n = 6$ eindeutig

identifizierbar; wissenschaftlicher Beruf: n = 6 eindeutig identifizierbar), sechs Personen machten hierzu keine Angabe.

14.2.2 Vorgehen

Es wurde eine Online-Umfrage (unipark, Questback GmbH) erstellt, die vom 15.12.2015 bis zum 11.08.2016 geöffnet war. Die Proband*innen wurden über E-Mail-Verteiler für Psychologie-Studierende (in Braunschweig und Osnabrück) sowie über soziale Medien akquiriert. Bei der Akquise wurde wie in der ersten Studie die Beschreibung „Studie zur Einschätzung von Personen" genutzt.

Nachdem die Teilnehmenden über die Freiwilligkeit, Anonymität und den Datenschutz aufgeklärt wurden, wurde zufällig eine von acht textbasierten Fallvignetten präsentiert, zu der die Proband*innen ihre persönliche Einschätzung geben sollten. Innerhalb der Fallvignetten, die auf denen von Zander et al. (2015) basierten, wurde eine Lehrkraft beschrieben, wobei das Geschlecht der Lehrkraft durch die Anrede (Herr, Frau), das Unterrichtsfach (Deutsch, Mathematik) und die Schulform (3. Klasse in der Grundschule, 8. Klasse im Gymnasium) zufällig variierten. Es handelte sich um ein 2 × 2 × 2-Design. Textbasierte Fallvignetten haben sich auch in bisher beschriebenen Studien zur sozialen Wahrnehmung und Kognition im Lehr-Lern-Kontext bewährt (u. a. Auwarter & Aruguete, 2008; Glock, 2016; Glock & Krolak-Schwerdt, 2013, 2014; Meltzer & McNulty, 2011; Müller et al., 2017; Zander et al., 2015). Im Folgenden ist die Vignette beispielhaft für die weibliche Deutschlehrkraft am Gymnasium präsentiert:

> Stellen Sie sich vor, Sie sind Schüler/in eines städtischen Gymnasiums. Mit dem Start in die achte Klasse bekommen Sie für den Deutschunterricht eine neue Fachlehrerin, Frau Beck. Frau Beck ist gerade auf die Schule gewechselt, nachdem sie ihr Referendariat in einem anderen Gymnasium abschloss und ein paar Jahre dort gearbeitet hat. In der ersten Deutschstunde des Schuljahres sehen Sie sie zum ersten Mal. Vor dem Unterricht ist es wie immer sehr laut, dann klingelt die Glocke. Frau Beck stellt sich am Anfang der Unterrichtsstunde vor und beginnt anschließend mit dem ersten Thema des Schuljahres.
>
> Bitte stellen Sie sich Frau Beck bildlich vor, wie sie den Unterricht durchführt. Nehmen Sie sich dafür etwas Zeit. Beantworten Sie anschließend spontan die folgende Frage. Es gibt keine richtigen oder falschen Antworten.

Anschließend wurden soziodemografische Daten erhoben und am Ende über das konkrete Anliegen der Befragung aufgeklärt sowie sich bei den Proband*innen für die Teilnahme bedankt. Die Daten wurden mittels IBM SPSS Statistics 25 und MPlus 8 ausgewertet.

14.2.3 Instrumente und Variablen

Als abhängige Variable wurde die Beurteilung der Kompetenz betrachtet und die Beurteilung der Wärme wurde als Kontrollvariable herangezogen (Fragebogen s. Zusatzmaterial).

Kompetenzbeurteilung: Die Einschätzung der Kompetenz wurde mittels der sieben Adjektive der ersten Studie erfasst (kompetent, fähig, effizient, intelligent, konkurrenzfähig, selbstbewusst, eigenständig), wobei nur auf die 4-Item-Lösung (kompetent, effizient, fähig, intelligent) zurückgegriffen wurde. Abweichend zur ersten Studie mit einer vierstufigen Skala sollten die Proband*innen ihre Zustimmung auf einer siebenstufigen Likert-Skala angeben (1 = „stimme gar nicht zu" bis 7 = „stimme vollkommen zu"), um differenziertere Antworten zu ermöglichen und somit die Varianz der Variablen zu erhöhen. Zusätzlich gab es eine „Keine Angabe"-Option. Die Reliabilität (Cronbachs Alpha) war mit $\alpha = .845$ zufriedenstellend.

Wärmebeurteilung: Die Einschätzung der Wärme wurde mit den gleichen sieben Adjektiven der ersten Studie (warmherzig, gutmütig, wohlwollend, sympathisch, offen, tolerant, vertrauenswürdig) und mit der gleichen siebenstufigen Likert-Skala (und zusätzlich „Keine Angabe") wie die Kompetenzbeurteilung erfasst. Die Items wurden zusammen mit denen der Kompetenzbeurteilung in einer gemeinsamen Adjektivliste zufällig variierend präsentiert. Es konnte für die ausgewählten vier Items der Wärmebeurteilung (warmherzig, gutmütig, wohlwollend, sympathisch) eine gute interne Konsistenz mit einem Cronbachs Alpha von $\alpha = .892$ erzielt werden. Die Wärme wurde auch in dieser Studie im Sinne des *stereotype content model* als Kontrollvariable herangezogen.

Die konfirmatorische Faktorenanalyse mit robuster Maximum-Likelihood-Schätzung (MLR) ergab im Vergleich zur ersten Studie schlechtere Fitmaße für zwei Faktoren mit je vier Items ($\chi^2/df = 2.982$; CFI = .942; TLI = .915; RMSEA = .092; SRMR = .062; Hu & Bentler, 1999; Weiber & Mühlhaus, 2014; Itemschwierigkeiten und Trennschärfen s. Zusatzmaterial). Eine einfaktorielle Lösung mit acht Items ergab keine Verbesserung der Fitmaße ($\chi^2/df = 10.290$; CFI = .735; TLI = .629; RMSEA = .191; SRMR = .124). Auch unter Berücksichtigung der ursprünglichen 14 Items in einem Faktor ($\chi^2/df = 6.274$; CFI = 717; TLI = .662; RMSEA = .144; SRMR = .133) oder in zwei Faktoren ($\chi^2/df = 2.533$; CFI = .917; TLI = .901; RMSEA = .078; SRMR = .083) konnten keine ausreichenden Fitmaße erzielt werden. Aufgrund der relativ kleinen Stichprobengröße, die der Faktorenanalyse zugrunde lag ($N < 250$; aufgrund von *missing data*), wurde ein stärkeres Gewicht auf den CFI und den SRMR gelegt (gemäß den Empfehlungen von Hu & Bentler, 1999; vgl. Weiber & Mühlhaus, 2014) und die 2-Faktoren-Lösung mit jeweils vier Items angenommen.

14.3 Ergebnisse

Die Kennwerte der Variablen sowie die Korrelationen sind in Tabelle 14.1 aufgelistet, die Kennwerte nach Gruppen können der Tabelle 14.2 entnommen werden. Da kaum Normalverteilungen vorlagen (auf Basis von Shapiro-Wilk-Tests) und die Quantil-Quantil-Diagramme optisch auffällig waren, wurde der Schwerpunkt der multivariaten Auswertung auf robuste und nicht-parametrische Berechnungen gelegt.

Die Hypothesen wurden in der bivariaten Betrachtung mit t-Tests überprüft. Es zeigten sich weder Geschlechts-Professions-Assoziation im Unterrichtsfach Deutsch ($t(104) = 0.539$, $p_1 = .296$; $d = 0.105$; $1–β = .134$) noch im Unterrichtsfach Mathematik ($t(122) = 0.456$, $p_1 = .325$; $d = 0.082$; $1–β = .117$). Auch bei der Kompetenzbeurteilung männlicher ($t(109) = 0.718$, $p_1 = .237$; $d = 0.137$; $1–β = .176$) und weiblicher Lehrkräfte in Abhängigkeit von der Schulform ($t(109.469) = 0.616$, $p_1 = .270$; $d = 0.113$; $1–β = .151$) gab es keine signifikanten Unterschiede. Dies bedeutet, dass sowohl Frauen als auch Männer als Lehrkräfte über die Schulformen hinweg gleich kompetent beschrieben wurden. Außerdem war der Unterschied zwischen der Kompetenzbeurteilung für weibliche und männliche Grundschullehrkräfte nicht signifikant ($t(106) = 0.063$, $p_1 = .475$; $d = 0.012$; $1–β = .057$). Insgesamt erhielten männliche und weibliche Lehrkräfte auch gleich hohe Kompetenzzuschreibungen ($t(106) = 0.063$, $p = .950$; $d = 0.008$; $1–β = .050$). Somit müssen auf Basis der bivariaten Betrachtungen alle aufgestellten Hypothesen abgelehnt werden.

Zur Überprüfung der Hypothesen unter Berücksichtigung der Wärmebeurteilung sollte eine Kovarianzanalyse berechnet werden. Die Voraussetzungen dafür lagen jedoch nicht vor. Zum einen gab es Verletzungen der Normalverteilungsannahme, wogegen varianzanalytische Verfahren allerdings robust sind (Harwell, 2003; vgl. Bortz & Schuster, 2010), außerdem gab es eine interaktive Wirkung einer Kontrollvariablen (Wärmebeurteilung) mit einem Faktor (Schulform) auf die Kompetenzbeurteilung. Letzteres kann zu einer Verzerrung der Ergebnisse führen (Huber et al., 2014). Deswegen wurde zur Hypothesenprüfung eine robuste Regressionsanalyse mit Bootstrap durchgeführt (vgl. Urban & Mayerl, 2018), in die die Variablen Geschlecht der Lehrkraft, Unterrichtsfach, Schulform und deren Interaktionsterme sowie die Wärmebeurteilung eingingen.

Tabelle 14.1 Kennwerte und Korrelationen der Variablen (Studie 2)

	N	*M*	*SD*	α	SW	Korrelationen[a] 1	2	3	4	5	6	7
*Proband*innen*												
1 Geschlecht[b]	252	1.82	0.39		.47**	-	−.17**	.13*	.10	−.04	.02	.04
2 Alter	253	24.94	5.89		.81**	−.19**	–	−.05	−.18**	−.00	−.01	−.01
Beurteilungen												
3 Kompetenzbeurteilung[c]	230	5.26	1.10	.85	.95**	−.04	.12	–	.42**	−.01	.11	.05
4 Wärmebeurteilung[c]	236	4.87	1.23	.89	.96**	−.11	.07	.51**	–	−.02	.04	.00
Bedingungen												
5 Geschlecht der LK[b]	254	1.52	0.50		.64*	.02	−.04	−.00	.01		−.01	−.03
6 Unterrichtsfach[d]	254	1.54	0.50		.63*	−.03	.02	.14*	.07	−.01	–	.12
7 Schulform[e]	254	1.52	0.50		.64*	−.05	.04	.06	.04	−.03	.12	–

Anmerkungen. α = Cronbachs Alpha; SW = Shapiro-Wilk-Test auf Normalverteilung; LK = Lehrkraft. [a] Korrelationen: $218 \leq N \leq 254$; über der Diagonalen: Korrelationen nach Spearman; unter der Diagonalen: Korrelationen nach Pearson. [b] Codierung: 1 = männlich, 2 = weiblich. [c] Mögliche Werte: 1–7. [d] Codierung: 1 = Deutsch, 2 = Mathematik. [e] Codierung: 1 = Grundschule, 2 = Gymnasium. * $p < .05$. ** $p < .01$.

Tabelle 14.2 Mittelwerte, Standardabweichungen und Normalverteilungsüberprüfungen der Skalen getrennt nach Gruppen (Studie 2)

		Kompetenzbeurteilung			Wärmebeurteilung		
Gruppe	*N*	*M*	*SD*	SW	*M*	*SD*	SW
Alle Gruppen	254	5.26	1.10	.95**	4.87	1.23	.96**
Männliche LK	123	5.26	1.14	.94**	4.86	1.32	.94**
Weibliche LK	131	5.25	1.07	.96**	4.89	1.14	.97**
Deutsch-LK	117	5.08	1.20	.93**	4.78	1.35	.94**
Mathematik-LK	137	5.40	1.00	.97**	4.96	1.11	.98*
Grundschul-LK	121	5.18	1.20	.95**	4.82	1.37	.94**
Gymnasial-LK	133	5.32	1.01	.95**	4.92	1.09	.97**
Männliche Deutsch-LK	61	5.02	1.11	.95*	4.87	1.27	.95*
Männliche Mathematik-LK	70	5.44	1.00	.96*	4.90	1.03	.98
Weibliche Deutsch-LK	56	5.15	1.29	.91**	4.69	1.43	.93**
Weibliche Mathematik-LK	67	5.36	1.00	.97	5.02	1.21	.96*
Männliche LK in der Grundschule	64	5.19	1.18	.96	4.78	1.29	.96
Männliche LK im Gymnasium	67	5.31	0.95	.94**	4.99	0.98	.96*
Weibliche LK in der Grundschule	57	5.18	1.23	.93**	4.87	1.45	.90**
Weibliche LK im Gymnasium	66	5.33	1.07	.94**	4.85	1.20	.97
Männliche Deutsch-LK in der Grundschule	31	4.87	1.17	.96	4.84	1.40	.94
Männliche Deutsch-LK im Gymnasium	30	5.19	1.04	.95	4.89	1.13	.92*
Männliche Mathematik-LK in der Grundschule	33	5.49	1.14	.94	4.72	1.19	.97
Männliche Mathematik-LK im Gymnasium	37	5.40	0.89	.95	5.05	0.86	.96
Weibliche Deutsch-LK in der Grundschule	32	5.05	1.40	.93*	4.59	1.59	.91*
Weibliche Deutsch-LK im Gymnasium	24	5.27	1.16	.80**	4.83	1.18	.94
Weibliche Mathematik-LK in der Grundschule	25	5.35	0.96	.95	5.24	1.18	.87**
Weibliche Mathematik-LK im Gymnasium	42	5.37	1.03	.95	4.87	1.22	.97

Anmerkungen. N: teils geringfügige Abweichung aufgrund fehlender Werte möglich. Mittelwerte von 1 bis 7 möglich. SW = Shapiro-Wilk-Test auf Normalverteilung; LK = Lehrkraft. * $p < .05$. ** $p < .01$.

Vorweg wurde eine Regressionsanalyse mit Bootstrap zur Überprüfung einer Eigengruppenbevorzugung (Balliet et al., 2014; DiDonato et al., 2011; Robbins & Krueger, 2005) auf Basis der Geschlechtszugehörigkeit durchgeführt. In diese gingen das Geschlecht der beschriebenen Lehrkräfte und das der Proband*innen als Regressoren ein. Dabei ergaben sich weder signifikante Haupteffekte (Geschlecht der Lehrkraft: $\beta = .138$, $p_{bt} = .649$; Geschlecht der Proband*innen: $\beta = -.031$, $p_{bt} = .058$) noch eine signifikante Interaktion ($\beta = .045$, $p_{bt} = .528$; s. Zusatzmaterial). Deswegen wurde das Geschlecht der Proband*innen nicht als Kontrollvariable in der explorativen multivariaten Regressionsanalyse berücksichtigt.

Tabelle 14.3 Ergebnisse der Regressionsanalyse mit der Kompetenzbeurteilung als Regressand (Studie 2)

			Bootstrap[a]		
			95 %–CI		
Variable	β/B[b]	p	*LL*	*UL*	p_{bt}
Konstante	0.000	1.000	−0.116	0.115	1.000
Haupteffekte					
Wärmebeurteilung	.510	< .001	0.370	0.639	< .001
Geschlecht der LK[c]	−.032	.586	−0.146	0.082	.582
Unterrichtsfach[d]	.088	.139	−0.027	0.203	.141
Schulform[e]	.039	.514	−0.076	0.152	.512
Interaktionen					
Geschlecht der LK x Unterrichtsfach	.088	.138	−0.026	0.205	.136
Geschlecht der LK x Schulform	−.050	.402	−0.167	0.063	.400
Unterrichtsfach x Schulform	−.068	.247	−0.181	0.043	.235
Geschlecht der LK x Unterrichtsfach x Schulform	−.015	.805	−0.128	0.097	.802
Modell	$R^2 = .289$, korrigiertes $R^2 = .261$; $F(8, 207) = 10.601$, $p < .001$				

Anmerkungen. Toleranz (Minimum) = .964. N = 218. LK = Lehrkraft. [a] Bootstrap: Perzentile; N = 10000. [b] Der standardisierte Regressionskoeffizient β entspricht aufgrund der vorherigen z-Transformierung aller Variablen dem unstandardisierten Koeffizienten B; für die Konstante und die Konfidenzintervalle ist B angegeben. [c] Codierung vor z-Transformation: 1 = männlich, 2 = weiblich. [d] Codierung vor z-Transformation: 1 = Deutsch, 2 = Mathematik. [e] Codierung vor z-Transformation: 1 = Grundschule, 2 = Gymnasium.

Die Ergebnisse der multivariaten Regressionsanalyse sind in Tabelle 14.3 dargestellt. Insgesamt konnten 26 % der Varianz der Kompetenzbeurteilung aufgeklärt werden. Dies lag im Haupteffekt der Wärmebeurteilung ($\beta = .510$, $p < .001$) begründet, der in die gleiche Richtung wies wie die Korrelation ($r_S = .423$, $p < .001$). Alle weiteren Regressoren sowie deren Interaktionen zeigten keinen signifikanten Zusammenhang mit der Kompetenzbeurteilung.

14.4 Zusammenfassung, Interpretationen und Limitationen

Mithilfe der zweiten Studie sollte eine zur ersten Untersuchung dieser Arbeit (s. Kapitel 13) zusätzliche Überprüfung von Geschlechts-Professions-Assoziationen bei Lehrkräften und somit von Substereotypisierungen vorgenommen werden. Außerdem sollte überprüft werden, ob Frauen im Grundschullehramt positiver resp. kompetenter beurteilt werden als Männer, da vermeintlich kommunale Eigenschaften bedeutsamer erscheinen.

Weder das Geschlecht der Lehrkraft noch das Unterrichtsfach und die Schulform konnten in der durchgeführten Studie mit textbasierten Fallvignetten bedeutsam die Kompetenzbeurteilung vorhersagen. Auch statistische Interaktionen blieben aus. Somit sind alle aufgestellten Hypothesen zu verwerfen. Es zeigten sich keine klassischen Geschlechts-Professions-Stereotype für Lehrkräfte – wie auch in den multivariaten Analysen der ersten Studie. Dies unterstützt die Interpretation der ersten Studie, dass lediglich wenige Unterrichtsfächer (z. B. Physik) mit männlichen oder weiblichen Lehrkräften assoziiert sind.

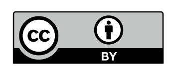
Außerdem scheint auf Basis der hier erzielten Ergebnissen der Erziehungsaspekt bei der Kompetenzbeurteilung von Lehrkräften keine Rolle zu spielen. Dies kann jedoch daran liegen, dass die Manipulation des Erziehungsaspekts innerhalb der Fallvignetten ungenügend war. Da in allen Fallvignetten eine Fachlehrkraft, auch in der Grundschule, beschrieben wird, kann die Beschreibung der Schulform – also Grundschule und Gymnasium, die unterschiedliche Anteile an Erziehungsaufgaben für Lehrkräfte innehaben – für die Beurteilung der Lehrkräfte irrelevant werden. Des Weiteren ist eine Fachlehrkraft in der Grundschule eher untypisch, was die Imagination der Situation für die Proband*innen erschwert haben bzw. Irritationen ausgelöst haben könnte.

Die Wärmebeurteilung stand in einem besonders hohen Zusammenhang mit der Kompetenzbeurteilung ($r_S = .423$, $p < .001$) und dies in einem höheren Ausmaß als in Studie 1 ($r_S = .343$, $p < .001$). Dies stützt die Interpretation, dass sympathisches und fürsorgliches Verhalten für Lehrkräfte einen besonderen

Kompetenzbereich darstellt. In Kompetenzmodellen für Lehrkräfte, z. B. im heuristischen Modell professioneller Handlungskompetenz COACTIV (Baumert & Kunter, 2006; Kunter et al., 2011), hat dieser Aspekt hingegen keine große explizite Bedeutsamkeit. Nichtsdestotrotz scheint es in der Lehrkräfteaus- und -weiterbildung lohnenswert zu sein, die Wärmedimension stärker zu berücksichtigen, wie z. B. in Klassenführungstrainings (z. B. Kiel et al., 2013; Marzano et al., 2003; vgl. Haag, 2018).

Die Limitationen gleichen denen der ersten Studie (s. Kapitel 13): Es wurde eine Gelegenheitsstichprobe herangezogen und die Proband*innen waren größtenteils Studierende und/oder aus dem psychosozialen Bereich. Dies begrenzt die Generalisierbarkeit (vgl. u. a. Bortz & Döring, 2016; Bühner, 2010). Außerdem wurden erneut ausschließlich kognitive Schemata von erwachsenen Proband*innen erfasst. Weitergehend waren die Items ausschließlich positiv gepolt.

Basierend auf den Ergebnissen kann insgesamt resümiert werden, dass es keine bedeutsamen Subgruppen von Lehrkräften gibt, die die Kompetenzbeurteilung in einem hohen Ausmaß beeinflussen. Auch bei Zander et al. (2015) ergibt sich lediglich ein kleiner Effekt für das Geschlecht der Lehrkraft auf die Beurteilung der Physiklehrfähigkeit. Nichtsdestotrotz ergeben sich Streuungen in der Beurteilung der Kompetenz. Auf welchen kognitiven Schemata oder situativen Einflüssen diese Streuungen beruhen, kann hier nicht abschließend geklärt werden und bedarf weiterer Forschung.





Studie 3 – Nonverbales Verhalten in Abhängigkeit von Kompetenzerwartungen

15

15.1 Hypothesen

Wenn Lehrpersonen eine Kompetenzerwartung an Lernende ausgebildet haben, wird diese über das Verhalten der Lehrpersonen an die Lernenden vermittelt (vgl. Brophy & Good, 1976; Nickel, 1985; Tauber, 1997). Schon Rosenthal und Jacobson (1971) argumentierten, dass sich Lehrkräfte, die eine*einen *intellectual bloomer* (schnellentwickelndes Kind) erwarteten, diesem Kind freundlicher und unterstützender begegneten. Diese Annahmen konnten meta-analytisch von Harris und Rosenthal (1985; Rosenthal, 1994; vgl. Ludwig, 2018) bestätigt werden, indem sie vier Faktoren als bedeutsam für die Vermittlung von Erwartungen ausmachten: die Atmosphäre (*climate*), das Feedback, den Input der Lehrpersonen und die Output-Möglichkeiten (s. Kapitel 6).

Es gibt nur wenige Befunde zur Mediation der Lernendenkompetenzerwartungen an Lehrpersonen (s. Kapitel 11): Jamieson et al. (1987) konnten ausmachen, dass sich Schüler*innen adäquater und weniger störend verhielten, wenn sie eine kompetente und motivierte Lehrperson erwarteten. Die Proband*innen in der Galatea-Bedingung in der ersten Studie von Feldman und Prohaska (1979) zeigten (marginal signifikant) ein zugewandteres Verhalten (*immediacy*; $d = 0.342$). Unter den von ihnen betrachteten Variablen waren für diesen Trend das Vorbeugen ($d = 0.974$) und der Blickkontakt ($d = 0.512$) verantwortlich, während die räumliche Nähe ($d = 0.187$) und die seitliche Orientierung ($d = -0.342$) kaum

---

Ergänzende Information Die elektronische Version dieses Kapitels enthält Zusatzmaterial, auf das über folgenden Link zugegriffen werden kann https://doi.org/10.1007/978-3-658-39444-8_15.

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_15

eine Rolle spielten. Ähnliches Verhalten in Abhängigkeit der Kompetenzerwartung konnte auch bei Lehrpersonen gefunden werden (Chaikin et al., 1974; vgl. Harris & Rosenthal, 1985).

Es ergibt sich also die Hypothese, dass sich Lernende mit hohen Kompetenzerwartungen (Galatea-Effekt) an Lehrpersonen im nonverbalen Verhalten von Lernenden mit geringen Kompetenzerwartungen an Lehrpersonen (Golem-Effekt) unterscheiden. Die bisherigen Befunde verweisen zwar auf ein zugewandteres Verhalten, wenn eine kompetente Person erwartet wird, aber die Studienergebnisse für das Lernendenverhalten sind rar und über 30 Jahre alt (Feldman & Prohaska, 1979; Jamieson et al., 1987), weswegen keine gerichteten Annahmen getroffen werden. Aus den genannten Befunden und Ableitungen ergeben sich folgende Hypothesen:

Hypothesenkomplex: Lernende, die eine kompetente Lehrperson erwarten, verhalten sich hinsichtlich (1.) des Blickkontakts, (2.) des Vorbeugens, (3.) des Abstützens des Kopfes auf dem Arm, (4.) des Verschränkens der Arme, (5.) des Lächelns, (6.) des Stirnrunzelns, (7.) des Nickens und (8.) des Kopfschüttelns anders als Lernende, die eine inkompetente Lehrperson erwarten.

Die Auswahl des relevanten nonverbalen Verhaltens basiert auf den Variablen, die Feldman und Prohaska (1979) untersuchten. Außerdem wurden ergänzende Verhaltensweisen betrachtet, um mehr Informationen zum Verhalten zu erhalten. So wurden das Abstützen des Kopfes auf dem Arm sowie das Verschränken der Arme als Gesten der Entspannung nach Mehrabian (1969) herangezogen sowie Nicken und Lächeln als Gesichtsausdruck (Chaikin et al., 1974; vgl. Mehrabian, 1969). Nach Mehrabian (1969) ist die Zugewandtheit (*immediacy*) ein Set von Verhaltensweisen (u. a. Berührung, Abstand, Vorbeugen, Blickkontakt), das Sympathie für andere widerspiegelt, wenn die Person nicht angespannt ist. Die Entspanntheit (*relaxation*) ist das zweite Set von Verhaltensweisen (u. a. asymmetrische Arm- und Beinhaltung, seitliches Abstützen), das Mehrabian (1969) beschreibt, und mit Status assoziiert: Je höher die Entspanntheit ist, desto höher ist der Status im Vergleich zu der*dem Interaktionspartner*in. Darüber hinaus wird, so Mehrabian (1969), die Einstellung zueinander in Interaktionen auch über Gesichtsausdrücke (u. a. Lächeln, Stirnrunzeln, Nicken, Kopfschütteln) kommuniziert.

Zudem soll überprüft werden, ob sich in Abhängigkeit von den Kompetenzerwartungen ein Lernzuwachs ergibt. Die Ergebnisse zum Lern- und

Leistungserfolg bzw. Wissenszuwachs in Abhängigkeit der Lernendenkompetenzerwartungen an Lehrpersonen sind uneindeutig: Die meisten Studien berichteten einen Lernvorteil (u. a. bei Adediwura & Tayo, 2007; Edwards et al., 2009; Jamieson et al., 1987; Reber et al., 2017), andere keinen (u. a. Lewandowski et al., 2012; Towler & Dipboye, 2006) und ein paar Studien konnten teilweise einen Lernerfolg ausmachen, z. B. bei bestimmten Tests (*privacy lesson*; Feldman & Prohaska, 1979) und in bestimmten Bedingungen (bei schlechten Vorträgen; Leventhal et al., 1977).

Hypothese 9: Lernende, die eine kompetente Lehrperson erwarten, erzielen ein besseres Ergebnis in einem Wissenstest als Lernende, die eine inkompetente Lehrperson erwarten.

Da in der Studie mit Videomaterial gearbeitet wurde (s. u.), wird die Attraktivitätsbeurteilung aufgrund des *„beauty is good" stereotype* (Dion et al., 1972; Eagly et al., 1991; Langlois et al., 2000) als Kontrollvariable herangezogen sowie auf Grundlage des *stereotype content model* (s. Kapitel 5) für die Wärmebeurteilung kontrolliert. Aufgrund der Zusammenhänge zwischen der Kompetenz-, Wärme- und Attraktivitätsbeurteilung könnte in der bivariaten Überprüfung die Wirkung der Kompetenzmanipulation (s. u.) überschätzt werden: So zeigte sich u. a. in Studie 1, die die Wahrnehmung von Lehrkräften mithilfe von Fallvignetten mit Porträts überprüfte, dass weibliche Lehrkräfte kompetenter beurteilt wurden ($d = 0.586$). Unter Kontrolle der Wärme- und Attraktivitätsbeurteilung war der Effekt jedoch kleiner ($\eta_p^2 = .057$ entspricht $d = 0.492$). Zudem soll in multivariaten Analysen ein Interaktionseffekt zwischen Lernendenkompetenzerwartungen und dem Geschlecht einer Lehrperson auf das nonverbale Verhalten der Lernenden und einen Wissenstest explorativ überprüft werden.

15.2 Methodik

15.2.1 Stichprobe

An der Untersuchung nahmen insgesamt 71 Personen teil, von denen acht ausgeschlossen wurden, weil sie an der Cover-Story (s. u.) zweifelten, und eine weitere, weil diese Person große Schwierigkeiten mit der deutschen Sprache aufwies. Somit ergab sich eine Stichprobe von 62 Proband*innen. Davon waren 50 weiblich und zwölf männlich. Das Durchschnittsalter betrug 24.8 Jahre ($SD = 8.94$; Spannweite: 18–53). Der Großteil der Teilnehmenden studierte

($n = 53$, davon 36 Psychologie und 9 Lehramt), neun Personen studierten nicht (mehr).

15.2.2 Vorgehen

Die experimentelle Laboruntersuchung erfolgte im Zeitraum vom 10.11.2016 bis zum 06.03.2017. Die Proband*innen wurden über Lehrveranstaltungen der Technischen Universität Braunschweig, E-Mail-Verteiler für Psychologie-Studierende der Technischen Universität Braunschweig, über das elektronische Schwarze Brett der Technischen Universität Braunschweig, über Aushänge in Supermärkten und über einen Sportverein akquiriert. Bei der Akquise wurde die Cover-Story verwendet, es handle sich um eine Überprüfung einer skandinavischen Lehrkräfte-Supervisionsmethode in Deutschland. Die Proband*innen erhielten entweder 5 € Aufwandsentschädigung, die Möglichkeit, an einer Gutscheinverlosung im Wert von 20 € teilzunehmen, oder eine Versuchspersonenstunde im Rahmen des Psychologie-Studiums.

Es wurden pro Durchgang bis zu vier Personen eingeladen, die zeitgleich im selben Raum teilnehmen konnten (Experimentalraum s. Abbildung 15.1). Durchschnittlich waren 1.82 Proband*innen ($SD = 1.02$) in insgesamt 39 Durchgängen präsent. Durchgeführt wurden alle Durchgänge der Untersuchung vom gleichen Experimentator (Autor der Arbeit), der nicht blind gegenüber den Versuchsbedingungen war. Der Experimentator orientierte sich dabei an einem Leitfaden.

Zu Beginn der Untersuchung wurde erneut die Cover-Story aufgegriffen und erklärt, dass die skandinavischen Länder in internationalen Schulvergleichsstudien sehr gut abschneiden würden, dort Supervisionen für Lehrkräfte stattfänden und dass das Ziel des Projekts die Überprüfung einer bestimmten skandinavischen Supervisionsmethode für Deutschland sei. Danach wurde auf die Freiwilligkeit, die (relative) Anonymität und den Datenschutz hingewiesen.

Anschließend erfolgte eine kurze Agenda für das Vorgehen. Dabei wurde behauptet, dass es zwei Teile gebe. Zum ersten Teil wurde gesagt, dass die Lehrkraft einen Vortrag halten wird, der via Live-Stream in den Raum übertragen wird. Weitergehend wurde behauptet, dass die Lehrkraft die Teilnehmenden sehen wird, wofür drei Kameras aus unterschiedlichen Winkeln aufgebaut wurden (links, rechts und unter dem Monitor; s. Abbildung. 15.1). Die Kameras, die das nonverbale Verhalten der Teilnehmenden erfassten, mussten gemäß der Cover-Story aus technischen Gründen Videos abspeichern, weswegen eine Einverständniserklärung der Proband*innen eingeholt wurde. Nach dem Einholen der



Abbildung. 15.1 Labor und Versuchsaufbau (Studie 3)

Einverständniserklärung erhielten die Proband*innen ein vermeintliches Zeugnis (Examenszeugnis und ein Arbeitszeugnis) der Lehrkraft (mit den Fächern Deutsch und Biologie). Über das Zeugnis wurde die Kompetenzerwartung manipuliert: Entweder erhielten die Teilnehmenden ein sehr gutes Zeugnis (mit einer Examensnote von 1.3 und einem guten Arbeitszeugnis; Galatea-Bedingung) oder ein befriedigendes Zeugnis (mit einer Examensnote von 3.3 und einem mäßigen Arbeitszeugnis; Golem-Bedingung; Zeugnisse s. Zusatzmaterial). Während des Lesens wurden die Kameras eingeschaltet. Nahm nur eine einzelne Person an einem Durchgang teil, wurden anstatt der drei möglichen nur zwei Kameras eingeschaltet.

Nach dem Lesen der Zeugnisse wurde der Vortrag der Lehrkraft angekündigt. Der Experimentator bat die Proband*innen, während des Vortrags nicht miteinander zu sprechen, und kündigte an, der Lehrkraft das Zeichen für den Start zu geben. Anschließend ging der Experimentator aus dem Raum, öffnete eine Tür in der Nähe und sagte, dass die Lehrkraft anfangen könne. Nach etwa 10–30 Sekunden Wartezeit startete der Experimentator von außen ein Video im Raum, in dem die Proband*innen saßen, das über einen Laptop auf einem Monitor abgespielt wurde. Die Teilnehmenden saßen in einem Abstand zwischen 2.50 bis 3.50 Meter Entfernung zum Monitor und der Experimentator war während des Vortrags nicht im Raum.

Es wurde eines von zwei unterschiedlichen Videos mit einer vermeintlichen Lehrkraft gezeigt. Untersuchungen zur Lehr-Lern-Interaktion mit Videos sind weitverbreitet (u. a. Edwards et al., 2009; Haefele, 1988; Lewandowski et al., 2012; Perry et al., 1979; Reavis, 1979; Towler & Dipboye, 2006). Die beiden Videos unterschieden sich inhaltlich nicht, in beiden ging es um das Thema Nägelkauen (Vortragstext in Anlehnung an Williams et al., 2006; s. Zusatzmaterial). Das Thema Nägelkauen wurde gewählt, weil davon ausgegangen wurde, dass hierzu wenig Vorwissen besteht. Die Videos unterschieden sich im Geschlecht der dargestellten Lehrkraft: Entweder wurde eine weibliche Lehrkraft gezeigt oder eine männliche (s. Abbildung 15.2). Dabei wurden beide Versionen von der gleichen Darstellerin eingesprochen, um persönliche Eigenschaften (z. B. Vortragsstil) konstant zu halten[1].

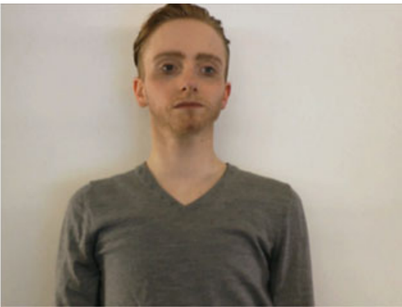



Abbildung 15.2 Konföderierte in der weiblichen (links) und der männlichen (rechts) Bedingung (Studie 3)

Das Video der weiblichen Lehrkraft-Bedingung dauerte 27:09 Minuten und das der männlichen Bedingung 25:56 Minuten. Die Videos wurden mit einer Panasonic-Kamera (HDC-SD 800) aufgenommen und mit Shotcut 16 bearbeitet (Farb- und Helligkeitskorrekturen; Weichzeichnung; 36 Bilder/Sekunde) sowie geschnitten (Porträt, etwa Zwerchfell aufwärts; Auflösung von 1200 × 900 Pixel). Außerdem wurde die Stimmhöhe mit Audacity 2.1 jeweils in einen „androgynen" Bereich verschoben. Dafür wurde ein circa einmütiger Ausschnitt der Audiospur zu Beginn der beiden Vortragsversionen bearbeitet, schrittweise in Seminaren des Instituts für Pädagogische Psychologie der Technischen Universität Braunschweig

[1] Eine vorangehende Erprobung der Untersuchung mit einer weiblichen und einem männlichen trainierten Konföderierten wurde aufgrund unterschiedlicher Vortragsweisen und Unglaubwürdigkeit, die von Proband*innen geäußert wurden, abgebrochen.

überprüft und so lange weiterbearbeitet, bis eine ungefähre Gleichverteilung des geschätzten Lehrkraft-Geschlechts erzielt wurde (insgesamt wurden acht Seminare besucht mit 137 teilnehmenden Studierenden). Dadurch wurde die Tonhöhe der männlichen Version um 3.0 Halbtöne reduziert (Bewertung im letzten besuchten Seminar: 39 % ($n = 7$) für weiblich, 61 % ($n = 11$) für männlich) und die der weiblichen Version um 3.2 Halbtöne (Bewertung im letzten besuchten Seminar: 42 % ($n = 8$) für weiblich, 58 % ($n = 11$) für männlich). Dies zeigt, dass die Darstellerin in der männlichen Version bereits tiefer sprach. Zusätzlich wurden der Pegel und die Lautstärke angepasst.

Nach Ende des Vortrags ging der Experimentator wieder in den Raum und teilte einen Fragebogen aus. Dieser beinhaltete die Einschätzung der vermeintlichen Lehrkraft, einen Wissenstest und Fragen zu soziodemografischen Daten. Danach wurden die Proband*innen über das Untersuchungsanliegen in vollem Umfang aufgeklärt und sie wurden gebeten, die genaue Fragestellung bis zum Ende der Untersuchung geheim zu halten. Außerdem wurden die Proband*innen gefragt, ob sie die Cover-Story glaubten, ob die aufgenommenen Videos von ihnen zum eigentlichen Zweck genutzt werden dürfen, und in der männlichen Bedingung, ob Zweifel am Geschlecht der Lehrkraft (erst offen, dann geschlossen) bestanden. Starke Zweifel an der Cover-Story äußerten acht Proband*innen und keine der teilnehmenden Personen zweifelte am Geschlecht der Lehrkraft in der männlichen Bedingung. Alle Proband*innen waren einverstanden, dass die Videos von ihnen im Rahmen der vorliegenden Arbeit genutzt werden dürfen.

Insgesamt handelte es sich also um ein 2×2-Design, bei dem einerseits die Kompetenzerwartungen über die Zeugnisse manipuliert wurden sowie andererseits das Geschlecht der Lehrkraft systematisch variierte. Zu Beginn der Untersuchung wurden die Bedingungen zufällig per Los bestimmt, im weiteren Verlauf wurde vermehrt auf eine gleiche Zellbesetzung Wert gelegt. Die Daten wurden mittels IBM SPSS Statistics 25 ausgewertet.

15.2.3 Instrumente und Variablen

Als abhängige Variablen wurden das nonverbale Verhalten der Proband*innen und das Wissen über Inhalte des Vortrags erfasst. Die Beurteilung der Kompetenz wurde als Manipulationsüberprüfung erfragt und zusätzlich wurden die Beurteilung der Wärme im Sinne des *stereotype content model* sowie die Beurteilung der Attraktivität aufgrund des *„beauty is good" stereotype* als Kontrollvariablen für multivariate Analysen herangezogen (Fragebogen s. Zusatzmaterial).

Kompetenzbeurteilung: Die Einschätzung der Kompetenz wurde mittels der vier Adjektive der ersten und zweiten Studie (kompetent, effizient, fähig, intelligent) auf einer siebenstufigen Likert-Skala (1 = „stimme gar nicht zu“ bis 7 = „stimme vollkommen zu“) erfasst. Die Reliabilität (Cronbachs Alpha) war mit $\alpha = .803$ zufriedenstellend.

Wärmebeurteilung: Die Einschätzung der Wärme wurde mit vier Adjektiven der ersten beiden Studien (warmherzig, gutmütig, wohlwollend, sympathisch) in der gleichen Liste wie die Kompetenzbeurteilung mit einer siebenstufigen Likert-Skala erfasst. Es konnte für die Wärmebeurteilung eine gute interne Konsistenz mit einem Cronbachs Alpha von $\alpha = .796$ erzielt werden.

Attraktivitätsbeurteilung: Die Einschätzung der Attraktivität wurde wie in der ersten Studie mit folgendem Item auf einer siebenstufigen Likert-Skala (1 = „sehr unattraktiv“ bis 7 = „sehr attraktiv“) erhoben: „Wie optisch attraktiv empfinden Sie die Lehrkraft?“

Auf Grundlage der Ergebnisse der Studie 1 und 2 sowie aufgrund der kleinen Stichprobengröße ($N = 62$) wurde auf faktoranalytische Überprüfungen verzichtet. Die internen Konsistenzen werden als ein Indiz für eine angemessene Passung betrachtet.

Nonverbales Verhalten: Aspekte des nonverbalen Verhaltens wurden auf Basis von Mehrabian (1969) und bereits existierender Studien (Chaikin et al., 1974; Feldman & Prohaska, 1979; Ho & Mitchell, 1982) ausgewählt: Vorgebeugt sein (Zeit), gerade sitzend (Zeit), zurückgelehnt sitzend (Zeit), Kopf auf dem Arm abgestützt (Zeit), Arme verschränkt (Zeit), Blickkontakt zum Monitor (Zeit), Weggucken (Häufigkeit), Lächeln (Zeit und Häufigkeit), Stirnrunzeln (Zeit und Häufigkeit), Kopfnicken (Zeit und Häufigkeit) und Kopfschütteln (Zeit und Häufigkeit; Manual s. Zusatzmaterial). Von den Videos wurden jeweils 164.4 Sekunden lange Ausschnitte kurz vor dem Ende des Vortrags (weibliche Bedingung: ab 26:57 Minuten; männliche Bedingung: ab 25:41 Minuten) ausgewählt, in denen in der männlichen und weiblichen Lehrkraft-Bedingung der gleiche Inhalt präsentiert wurde. Das Ende wurde herangezogen, weil davon ausgegangen wurde, dass sich zu diesem Zeitpunkt die Proband*innen an die Situation (z. B. die Kameras) gewöhnt hatten. Außerdem zeigte sich in der Studie von Herrell (1971), dass nur zum Ende Unterschiede bezüglich der (Wärmeerwartungs-)Manipulation beobachtbar waren. Die Videoausschnitte wurden dabei ohne Ton und mit 25 Bildern/Sekunde gespeichert. Die Bearbeitung und das Schneiden der Videos erfolgten mit Shotcut 16.

Das nonverbale Verhalten wurde in einem ersten Schritt von zwei trainierten Auswerterinnen, die blind gegenüber der Fragestellung waren, beurteilt.

Dafür wurde das Programm Avidemux 2.6 verwendet. Im Gegensatz zu Shotcut ist Avidemux ressourcenschonender und einfacher in der Anwendung für die Auswertung. Aufgrund des Formats von 25 Bildern/Sekunde entsprach die Zeit von einem Bild zum nächsten 0.04 Sekunden. Im Anschluss daran wurden die Übereinstimmungen mittels Korrelationen und in Anlehnung an die Bland–Altman-Methode (Differenz der Mittelwerte darf nicht signifikant von null verschieden sein; vgl. Grouven et al., 2007) berechnet. Dabei zeichnete sich ab, dass einige Variablen kaum übereinstimmend beurteilt wurden (gerade sitzend, zurückgelehnt sitzend, Lächeln, Stirnrunzeln) und andere Verhaltensweisen kaum auftraten (Kopfnicken, Kopfschütteln), die dann von der weiteren Betrachtung ausgeschlossen wurden. Danach wurden in moderierten Gesprächen größere Abweichungen in den Auswertungen (Zeit: über 5 Sekunden; Häufigkeit: über 1 Mal) der übrigen Variablen diskutiert und ggf. angepasst. Am Ende konnten fünf Variablen mit hoher Übereinstimmung herangezogen werden: Vorgebeugt sein (Zeit), Kopf auf dem Arm abgestützt (Zeit), Arme verschränkt (Zeit), Blickkontakt (Zeit) und Weggucken (Häufigkeit; s. Zusatzmaterial, inkl. Interraterreliabilität). Als Werte für das Verhalten wurden die Mittelwerte der beiden Auswerterinnen herangezogen.

Wissen: Das Wissen wurde mit zehn Fragen im Single Choice-Format erfasst. Von diesen zehn Fragen wurden zwei ausgeschlossen, da diese nicht exakt über den Vortrag beantwortbar waren bzw. nicht eindeutig formuliert wurden und ein weiteres Item aufgrund einer psychometrisch zu hohen Itemschwierigkeit (von mehr als 95 % richtig beantwortet; Kelava & Moosbrugger, 2020; Itemanalyse s. Zusatzmaterial). Somit war ein Testergebnis von 0 bis 7 möglich. Die Items hingen jedoch nicht bedeutsam untereinander zusammen (maximales $|r| = .228$, $p = .075$; Cronbachs Alpha: $\alpha = .244$).

Weitere Variablen: Zusätzlich zu den zentralen Variablen wurden soziodemografische Daten erfasst. Außerdem wurden die Uhrzeit des geplanten Beginns des Untersuchungsdurchgangs und bisheriges eigenes Nägelkauen zur Kontrolle erfasst. Die Uhrzeit wurde berücksichtigt, weil sie (bzw. der zirkadiane Rhythmus) die kognitive Leistung und Konzentration beeinflussen kann (u. a. Blatter & Cajochen, 2007). Da das Untersuchungsmaterial das Thema Nägelkauen aufgriff, wurde außerdem das Nägelkauen der Proband*innen erfragt.

15.3 Ergebnisse

Die Kennwerte der Variablen sind in Tabelle 15.1, die Kennwerte nach Gruppen in Tabelle 15.2 sowie die Korrelationen in der Tabelle 15.3 aufgelistet

Tabelle 15.1 Kennwerte der Variablen (Studie 3)

		Spannweite					
	N	Minimum	Maximum	*M*	*SD*	α/ICC[a]	SW
Geschlecht der Proband*innen[b]	62	1	2	1.81	0.40		.48**
Alter der Proband*innen	62	18	53	24.79	8.94		.72**
Nägelkauen (Gegenwart und/oder Vergangenheit)[c]	60	0	1	0.33	0.48		.60**
Kompetenzmanipulation[d]	62	1	2	1.48	0.50		.64**
Geschlecht der Lehrkraft[b]	62	1	2	1.48	0.50		.64**
Kompetenzbeurteilung[e]	61	1.50	7.00	5.32	1.23	.80	.93**
Wärmebeurteilung[e]	62	2.00	7.00	4.94	1.04	.80	.98
Attraktivitätsbeurteilung[e]	62	2.00	6.00	4.19	1.10		.91**
Blickkontakt[f]	62	70.37	164.40	140.68	19.90	1.00*	.91**
Weggucken	62	0.0	49.0	12.90	10.28	1.00*	.89**
Vorgebeugt[f]	62	0.00	164.40	20.69	51.01	1.00*	.44**
Kopf abgestützt[f]	62	0.00	164.40	41.66	60.46	1.00*	.70**
Arme verschränkt[f]	62	0.00	164.40	26.11	54.85	1.00*	.52**
Wissenstest[g]	62	3	7	5.74	1.13	.24	.87**
Uhrzeit	62	9.00	19.00	13.51	3.24		.91**

Anmerkungen. SW = Shapiro–Wilk-Test auf Normalverteilung. [a] Cronbachs Alpha (α) für Fragebogenmaße und Intraklassenkorrelation (ICC; Two-Way-Random, absolute Übereinstimmung, Durchschnitt) der Auswerterinnen für nonverbales Verhalten. [b] Codierung: 1 = männlich, 2 = weiblich. [c] Codierung: 0 = nein, 1 = ja. [d] Codierung: 1 = inkompetent, 2 = kompetent. [e] Mögliche Werte: 1–7. [f] Einheit: Sekunden (max. 164.40). [g] Mögliche Werte: 0–7. * $p < .05$. ** $p < .01$.

(Itemanalysen s. Zusatzmaterial). Da kaum Normalverteilungen vorlagen (Shapiro–Wilk-Test) und die Quantil-Quantil-Diagramme auffällig waren, wurde der Schwerpunkt der Auswertung auf robuste und nicht-parametrische Berechnungen gelegt.

Für die bivariaten Analysen wurden Mann–Whitney-U-Tests herangezogen, da wenig Proband*innen teilnahmen und die Variablen häufig nicht normalverteilt sind (s. Tabelle 15.2). Die Ergebnisse der Mann–Whitney-U-Tests sind in Tabelle 15.4 dargestellt. Es gab signifikante Unterschiede in den Bedingungen der Kompetenzmanipulation beim Blickkontakt, sowohl Dauer ($U = 339.0$,

Tabelle 15.2 Mittelwerte, Standardabweichungen und Normalverteilungsüberprüfungen der Skalen getrennt nach Gruppen (Studie 3)

Gruppe	*N*	*M*	*SD*	SW	*M*	*SD*	SW	*M*	*SD*	SW
		Kompetenzbeurteilung[a]			Wärmebeurteilung[a]			Attraktivitätsbeurteilung[a]		
Alle Gruppen	62	5.32	1.23	.93**	4.94	1.04	.98	4.19	1.10	.91**
Männliche LK	32	5.49	1.07	.93*	5.17	1.18	.96	3.78	1.13	.92*
Weibliche LK	30	5.14	1.37	.93	4.69	0.82	.97	4.63	0.89	.86**
Inkompetente LK	32	5.14	1.29	.93*	4.75	0.91	.96	3.84	1.08	.88**
Kompetente LK	30	5.52	1.15	.86**	5.14	1.15	.97	4.57	1.01	.90**
MI	16	5.30	1.28	.90	4.94	1.07	.92	3.31	1.08	.86*
MK	16	5.70	0.77	.93	5.41	1.27	.92	4.25	1.00	.90
WI	16	4.98	1.31	.94	4.56	0.70	.98	4.38	0.81	.79**
WK	14	5.32	1.45	.86*	4.84	0.94	.93	4.93	0.92	.87*
		Blickkontakt[b]			Weggucken			Vorgebeugt[b]		
Alle Gruppen	62	140.68	19.90	.91**	12.90	10.28	.89**	20.69	51.01	.44**
Männliche LK	32	138.16	22.02	.90**	13.05	9.55	.90**	19.23	49.51	.43**
Weibliche LK	30	143.36	17.33	.92*	12.75	11.17	.87**	22.24	53.36	.46**
Inkompetente LK	32	145.63	17.14	.88**	9.72	7.17	.93*	26.32	55.83	.52**
Kompetente LK	30	135.39	21.52	.92*	16.30	12.01	.89**	14.67	45.49	.35**
MI	16	141.98	19.59	.88*	10.09	6.92	.94	28.18	56.64	.56**
MK	16	134.33	24.23	.91	16.00	11.04	.89	10.28	41.10	.27**
WI	16	149.29	13.97	.87*	9.34	7.62	.89	24.47	56.81	.48**
WK	14	136.59	18.77	.94	16.64	13.45	.88	19.70	51.14	.45**

(Fortsetzung)

Tabelle 15.2 (Fortsetzung)

Gruppe	*N*	*M*	*SD*	SW	*M*	*SD*	SW	*M*	*SD*	SW
		Kopf abgestützt[b]			Arme verschränkt[b]			Wissenstest[a]		
Alle Gruppen	62	41.66	60.46	.70**	26.11	54.85	.52**	5.74	1.13	.87**
Männliche LK	32	36.59	56.77	.67**	22.97	52.98	.48**	5.84	1.30	.81**
Weibliche LK	30	47.07	64.70	.71**	29.47	57.48	.56**	5.63	0.93	.88**
Inkompetente LK	32	57.31	65.93	.76**	16.08	40.67	.46**	5.62	1.10	.90**
Kompetente LK	30	24.96	49.85	.57**	36.82	65.80	.57**	5.87	1.17	.85**
MI	16	56.10	66.29	.76**	15.40	41.29	.45**	5.50	1.21	.91
MK	16	17.07	38.12	.54**	30.53	63.04	.52**	6.19	1.33	.66**
WI	16	58.52	67.71	.78**	44.01	70.48	.47**	5.75	1.00	.88*
WK	14	33.98	60.85	.62**	16.75	41.38	.63**	5.50	0.85	.86*

Anmerkungen. N: teils geringfügige Abweichung aufgrund fehlender Werte möglich. SW: Shapiro–Wilk-Test auf Normalverteilung; LK = Lehrkraft; MI = männliche, inkompetente Lehrkraft; MK = männliche, kompetente Lehrkraft; WI = weibliche, inkompetente Lehrkraft; WK = weibliche, kompetente Lehrkraft. [a] Mögliche Werte: 1–7. [b] Mögliche Werte: 0.00–164.40 s. * $p < .05$. ** $p < .01$.

Tabelle 15.3 Korrelationen der Variablen (Studie 3)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Geschlecht der Pbn[a]	–	−.38**	−.16	−.34**	−.18	−.01	.02	−.23	.21	−.19	.01	.06	−.12	−.09	−.16
2	Alter der Pbn	.44**	–	.10	.31*	−.01	.10	−.03	.25	−.14	−.02	−.07	−.10	.12	.09	.38**
3	Nägelkauen der Pbn[b]	−.16	.11	–	−.02	.02	−.06	.01	−.01	−.03	.00	−.08	−.12	.20	−.08	.18
4	Kompetenzmanipulation[c]	−.34**	.27*	−.02	–	−.03	.14	.21	.35**	−.25*	.29*	−.18	−.24	.13	.12	.09
5	Geschlecht der LK[a]	−.18	−.01	.02	−.03	–	−.12	−.28*	.38**	.10	−.06	−.02	.09	.09	−.14	.11
6	Kompetenzbeurteilung	−.01	.11	−.11	.16	−.14	–	.51**	.19	.03	−.03	−.26*	.09	−.09	−.12	−.08
7	Wärmebeurteilung	.04	.12	.01	.19	.23	.53**	–	.25	−.03	.19	−.15	.04	−.19	.04	−.12
8	Attraktivitätsbeurteilung	−.21	.27*	.02	.33**	.39**	.16	.31*	–	.02	−.01	−.11	.06	.14	.16	.11
9	Blickkontakt	.18	−.16	−.00	−.26*	.13	−.02	−.03	−.04	–	−.79**	−.20	−.08	−.10	−.03	−.27*
10	Weggucken	−.17	.09	.10	.32*	−.02	.08	.15	.05	−.63**	–	.32*	.10	.03	.14	.14
11	Vorgebeugt	−.02	−.14	−.15	−.12	.03	−.16	−.18	−.03	−.12	.29*	–	.36**	.00	−.13	−.09
12	Kopf abgestützt	.10	−.06	−.08	−.27*	.09	.14	.04	.01	.04	−.04	.42**	–	−.12	−.06	−.10
13	Arme verschränkt	−.19	−.08	.22	.19	.06	−.10	−.21	.12	.03	.06	−.04	−.24	–	.25	.11
14	Wissenstest	−.09	.07	−.05	.11	−.09	−.08	.06	.16	.03	−.01	−.02	−.07	.17	–	−.23
15	Uhrzeit	.16	.39**	.19	.07	.12	−.10	−.15	.11	−.26*	.20	−.08	−.12	.13	−.30*	–

Anmerkungen. Über der Diagonalen: Korrelationen nach Spearman; unter der Diagonalen: Korrelationen nach Pearson. $59 \leq N \leq 62$. Pbn = Proband*innen; LK = Lehrkraft. [a] Codierung: 1 = männlich, 2 = weiblich. [b] Codierung: 0 = nein, 1 = ja. [c] Codierung: 1 = inkompetent, 2 = kompetent. * $p < .05$. ** $p < .01$

$p = .047$; $d = 0.521$; $1-\beta = .504$; Adjustierung[2]: $p = .047$) als auch Häufigkeit des Wegguckens ($U = 321.0$, $p = .025$; $d = 0.594$; $1-\beta = .612$; Adjustierung: $p = .047$). D. h., Teilnehmende in der Bedingung mit hohen Kompetenzerwartungen zeigten weniger Blickkontakt zum Monitor. Ein marginal signifikanter Effekt ergab sich beim Abstützen des Kopfes ($U = 358.5$, $p = .059$; $d = 0.495$; $1-\beta = .464$), wobei Proband*innen in der Galatea-Bedingung weniger lang den Kopf abstützten. Die Kompetenzmanipulation hatte keinen signifikanten Effekt auf die anderen Variablen des nonverbalen Verhaltens oder auf das Ergebnis im Wissenstest. Damit kann nur die Hypothese für den Blickkontakt angenommen werden.

Tabelle 15.4 Ergebnisse der Mann–Whitney-U-Tests für das nonverbale Verhalten und den Wissenstest in Abhängigkeit der Kompetenzmanipulation (Studie 3)

			p	
	U	*Z*	asymptotisch	exakt
Blickkontakt	339.0	−1.986	.047	.047
Weggucken	321.0	−2.241	.025	.025
Vorgebeugt	412.0	−1.440	.150	.175
Kopf abgestützt	358.5	−1.890	.059	.059
Arme verschränkt	425.5	−0.998	.318	.324
Wissenstest	415.0	−0.951	.342 ($p_1 = .171$)	.349 ($p_1 = .174$)

Anmerkungen. $N = 62$.

Zur Überprüfung der Hypothesen unter Berücksichtigung von Kontrollvariablen wurden robuste Regressionsanalysen mit Bootstrap (vgl. Urban & Mayerl, 2018) berechnet. Aufgrund der Verteilung der Variablen Vorgebeugt sein (von 11 Personen gezeigt), Kopf abgestützt (von 27 Personen gezeigt) und Arme verschränkt (von 16 Personen gezeigt) wurden für diese binär logistische Regressionsanalysen berechnet (Verhalten wurde gezeigt oder nicht gezeigt). Anzumerken ist, dass für diese logistischen Regressionsanalysen die geforderte Fallanzahl von mindestens 25 je Gruppe (aufseiten der abhängigen Variablen; vgl. Backhaus et al., 2016) unterschritten wurde. Für die Dauer des Blickkontakts, die Häufigkeit des Wegguckens und das Ergebnis im Wissenstest wurden lineare Regressionen mit Bootstrap berechnet. In allen Regressionsanalysen waren die

[2] Signifikanzadjustierung nach Benjamini und Hochberg (1995) mit dem Online-Tool von Hemmerich (2016).

Tabelle 15.5 Ergebnisse der linearen Regressionsanalysen (Studie 3)

Variable	Blickkontakt					Weggucken					Wissenstest				
			Bootstrap[a]					Bootstrap[a]					Bootstrap[a]		
			95 %–CI					95 %–CI					95 %–CI		
	β/B[b]	p	LL	UL	p_{bt}	β/B[b]	p	LL	UL	p_{bt}	β/B[b]	p	LL	UL	p_{bt}
Konstante	−0.000	1.000	−0.237	0.251	1.000	−0.000	1.000	−0.238	0.247	1.000	0.000	1.000	−0.239	0.239	1.000
Kontrollvariablen															
Alter der Pbn	.013	.927	−0.276	0.300	.927	−.085	.559	−0.357	0.159	.491	−.031	.826	−0.322	0.294	.839
Uhrzeit	−.267	.062	−0.522	0.016	.056	.245	.082	−0.003	0.531	.066	−.327	.020	−0.674	−0.023	.054
Wärme[c]	.008	.955	−0.360	0.268	.954	.200	.172	−0.033	0.499	.131	−.142	.322	−0.387	0.103	.252
Attraktivität[c]	.000	.999	−0.394	0.375	.999	−.162	.328	−0.492	0.116	.284	.288	.080	−0.040	0.565	.066
Regressoren															
1 Kompetenzmanipulation	−.240	.086	−0.504	0.018	.079	.346	.014	0.094	0.627	.015	.064	.630	−0.194	0.305	.609
2 Geschlecht LK[d]	.157	.305	−0.182	0.492	.378	.077	.611	−0.196	0.366	.591	−.199	.182	−0.484	0.097	.182
Interaktion 1 × 2	−.066	.605	−0.307	0.172	.596	.035	.777	−0.206	0.299	.780	−.190	.126	−0.432	0.048	.131
Modell	$R^2 = .155$, korrigiertes $R^2 = .045$; $F(7, 54) = 1.410$, $p = .220$					$R^2 = .176$, korrigiertes $R^2 = .069$; $F(7, 54) = 1.643$, $p = .143$					$R^2 = .205$, korrigiertes $R^2 = .102$; $F(7, 54) = 1.785$, $p = .074$				

Anmerkungen. Toleranz (Minimum) = .568. $N = 62$. Pbn = Proband*innen; LK = Lehrkraft. [a] Bootstrap: Perzentile; $N = 10000$. [b] Der standardisierte Regressionskoeffizient β entspricht aufgrund der vorherigen z-Transformierung aller Variablen dem unstandardisierten Koeffizienten *B*; für die Konstanten und die Konfidenzintervalle ist *B* angegeben. [c] Subjektive Beurteilung der gezeigten Lehrkraft. [d] Codierung vor z-Transformation: 1 = männlich, 2 = weiblich.

Tabelle 15.6 Ergebnisse der binär logistischen Regressionsanalysen (Studie 3)

Variable	Vorgebeugt					Kopf abgestützt					Arme verschränkt				
			95 %–CI					95 %–CI					95 %–CI		
	OR	*p*	*LL*	*UL*	p_{bt}[a]	OR	*p*	*LL*	*UL*	p_{bt}[a]	OR	*p*	*LL*	*UL*	p_{bt}[a]
Konstante	0.131	<.001			<.001	0.751	.285			.328	0.282	<.001			<.001
Kontrollvariablen															
Alter der Pbn	0.507	.330	0.129	1.992	.206	1.174	.605	0.639	2.160	.611	0.726	.437	0.324	1.627	.344
Uhrzeit	0.755	.499	0.335	1.703	.478	0.925	.796	0.514	1.665	.786	1.045	.899	0.530	2.058	.901
Wärme[b]	0.452	.104	0.173	1.178	.122	1.278	.458	0.668	2.445	.495	0.369	.014	0.166	0.819	.015
Attraktivität[b]	1.005	.992	0.376	2.683	.991	1.281	.496	0.628	2.615	.563	2.283	.070	0.935	5.575	.131
Regressoren															
1 Kompetenzmanipulation	0.627	.280	0.270	1.461	.121	0.550	.054	0.300	1.010	.074	1.202	.589	0.617	2.342	.581
2 Geschlecht LK[c]	1.006	.990	0.400	2.530	.517	1.093	.787	0.573	2.086	.801	0.746	.453	0.346	1.605	.502
Interaktion 1 x 2	1.643	.237	0.721	3.745	.089	1.017	.951	0.595	1.737	.942	1.156	.659	0.607	2.201	.627
Modell	Cox & Snell $R^2 = .127$, Nagelkerkes $R^2 = .210$; Richtige = 83.9 %; $\chi^2(7) = 8.444$, $p = .295$					Cox & Snell $R^2 = .079$, Nagelkerkes $R^2 = .105$; Richtige = 59.7 %; $\chi^2(7) = 5.071$, $p = .651$					Cox & Snell $R^2 = .139$, Nagelkerkes $R^2 = .204$; Richtige = 77.4 %; $\chi^2(7) = 9.268$, $p = .234$				

Anmerkungen. Es gingen z-standardisierte Werte ein. $N = 62$. Pbn = Proband*innen; LK = Lehrkraft. [a] Bootstrap: Perzentile; $N = 10000$, für Vorgebeugt 9998. [b] Subjektive Beurteilung der gezeigten Lehrkraft. [c] Codierung vor z-Transformation: 1 = männlich, 2 = weiblich.

Kompetenzmanipulation und das Geschlecht der Lehrkraft sowie deren Interaktion die Regressoren. Die Attraktivitäts- („*beauty is good*“ *stereotype*) und Wärmebeurteilung (*stereotype content model*) sowie die Uhrzeit (aufgrund des zirkadianen Rhythmus; u. a. Blatter & Cajochen, 2007) wurden als Kontrollvariablen berücksichtigt.

Es zeigten sich signifikante Korrelationen zwischen dem Alter ($r_S = .308$, $p = .016$) und dem Geschlecht der Proband*innen ($r_S = -.343$, $p = .006$) mit der Kompetenzmanipulation, sodass die älteren häufiger in der Galatea-Bedingung waren und die weiblichen Teilnehmenden häufiger in der Golem-Bedingung. In den Korrelationsanalysen zeigte sich auch, dass das Alter und das Geschlecht der Proband*innen zusammenhingen ($r_S = -.375$, $p = .003$), dass also die teilnehmenden Frauen jünger waren als die Männer. Außerdem nahmen die älteren Proband*innen zu einer späteren Tageszeit teil als die jüngeren ($r_S = .375$, $p = .003$). Deswegen wurde das Alter der Proband*innen als zusätzliche Kontrollvariable in den Regressionsanalysen berücksichtigt. Das Geschlecht wurde jedoch aufgrund der kleinen männlichen Stichprobe ($n = 12$) nicht als Kontrollvariable berücksichtigt. Es gab keine signifikanten Zusammenhänge mit dem eigenen regelmäßigen Nägelkauen (in der Vergangenheit oder in der Gegenwart) mit einer der abhängigen Variablen, weswegen das Nägelkauen nicht als Kontrollvariable in das Modell aufgenommen wurde.

Regressionsanalytisch (s. Tabelle 15.5 und Tabelle 15.6) zeigte sich, dass insgesamt wenig Varianz des nonverbalen Verhaltens aufgeklärt werden konnte (korrigiertes $R^2 \leq .102$; Nagelkerkes $R^2 \leq .210$). Ein signifikanter Effekt für die Kompetenzmanipulation ergab sich beim Weggucken (Häufigkeit; $\beta = .346$, $p_{bt} = .015$; Adjustierung: $p = .030$) sowie marginal signifikante Effekte beim Blickkontakt (Zeit; $\beta = -.240$, $p_{bt} = .079$; Adjustierung: $p = .079$) und beim Abstützen des Kopfes ($OR = 0.550$, $p_{bt} = .074$). Unter der Berücksichtigung der kleinen Stichprobe ($N = 62$) können also bei Annahme von marginal signifikanten Ergebnissen (vgl. Hager, 2004) die Hypothesen 1 und 3 unter Vorbehalt angenommen werden.

Zusätzlich zu den hypothesenprüfenden, multivariaten Regressionsanalysen wurde explorativ die Beurteilung der Kompetenz als Regressand betrachtet, da diese sich zwischen der Galatea- und der Golem-Bedingung nicht unterschied ($U = 389.0$, $p_{exakt} = .281$), was jedoch zu erwarten gewesen wäre. Es zeigte sich, dass nur die Wärmebeurteilung ein signifikanter Regressor für die Kompetenzbeurteilung war ($\beta = .509$, $p_{bt} < .001$), die Kompetenzmanipulation aber nicht ($\beta = .509$, $p_{1-bt} = .332$; s. Zusatzmaterial).

15.4 Zusammenfassung, Interpretationen und Limitationen

Die dritte Studie widmete sich dem Einfluss der Kompetenzerwartungen auf das nonverbale Verhalten. Mit einem experimentellen Design, in dem die Kompetenzerwartungen systematisch manipuliert wurden, zeigte sich, dass positive Kompetenzerwartungen (Galatea-Bedingung) mit weniger Blickkontakt – sowohl zeitlich kürzer als auch häufigerem Abbruch des Blickkontakts – und (marginal signifikant) weniger langem Abstützen des Kopfes einhergehen. Dies weist auf unterschiedliches Verhalten in Abhängigkeit der Kompetenzerwartung hin, widerspricht jedoch die bisherigen Befunde, die zeigten, dass Personen sich zugewandter verhalten, wenn sie eine kompetente Lehrperson erwarten, im Vergleich zu Personen, die eine inkompetente Lehrperson erwarten.

Ein zugewandteres Verhalten Lernender bei höheren Lernendenkompetenzerwartungen an Lehrpersonen konnte von Feldman und Prohaska (1979) ausgemacht werden. Für die Dauer des Blickkontakts ergab sich ein mittlerer Effekt ($d = -0.521$), bei Feldman und Prohaska (1979) zeigte sich ein ähnlicher Effekt ($d = .512$), allerdings in entgegengesetzter Richtung (mehr Blickkontakt bei der Galatea-Bedingung). Die Effektstärke bezüglich des Vorbeugens fällt auch entgegengesetzt und zusätzlich geringer aus ($d = -0.372$) als der Unterschied, der bei Feldman und Prohaska (1979) beschrieben ist ($d = 0.974$), was auch auf das unterschiedliche Erhebungsverfahren zurückführbar sein kann. So erfassten Feldman und Prohaska (1979) den genauen Grad des Vor- und des Zurückbeugens. Darüber hinaus gab es bei Feldman und Prohaska (1979) dyadische Interaktionen, in der hier durchgeführten Studie haben ein bis vier Proband*innen pro Untersuchungsdurchgang eine videografierte Lehrperson erlebt.

Die teilweise widersprechenden Ergebnisse zum nonverbalen Verhalten könnten mit zwei unterschiedlichen Motivationszuständen resp. mittels zweier „Langeweileprofile“ erklärt werden: So lag bei Feldman und Prohaska (1979) eine dyadische Face-to-Face-Interaktion vor, die vermutlich aktivierender ist als das Betrachten eines Monitors wie in der hier durchgeführten Studie. Es kann vermutet werden, dass in einem aktivierenden Setting Lernende ihr Desinteresse, das durch die Kompetenzerwartungen beeinflusst sein kann, durch Zurücklehnen, seitliches Eindrehen und Weggucken signalisieren. Bei einem monotonen Vortrag könnte sich Desinteresse der Zuhörenden in ein „einsinkendes“ Verhalten (z. B. Kopf auf dem Arm abgestützt) abbilden. Das Abstützen des Kopfes auf den Armen kann als entspannte Haltung (*relaxation*) interpretiert werden, was nach der Theorie von Mehrabian (1969) mit Status assoziiert ist. Personen mit höherem Status zeigen nach Mehrabian (1969) eine entspanntere Körperhaltung, sodass im

Fall von Lernenden eine entspanntere Körperhaltung so interpretiert werden kann, dass die Lernenden die jeweilige Lehrperson nicht als Expert*in wahrnehmen und somit nicht ernst nehmen. Sowohl das aktivierende Face-to-Face- als auch das monotone Vortragssetting haben dann gemeinsam, dass Lernende in der Galatea-Bedingung eine Haltung mit mehr Muskeltonus zeigen, nur wird beim Vortrag weitere Stimulation für kurze Zeit außerhalb des Settings gesucht (Weggucken), was wiederum zu einer Reduzierung von Müdigkeit führen könnte.

Das Geschlecht der Lehrperson hat in dieser Studie keinen signifikanten Effekt auf das nonverbale Verhalten und auch keinen bedeutsamen Effekt auf den Wissenstest. Dies kann auf eine geringe Bedeutung des Lehrpersonengeschlechts in der Lehr-Lern-Interaktion hinweisen (Basow et al., 2013; Bavishi et al., 2010; Goebel & Cashen, 1979; Riniolo et al., 2006; s. auch Studie 1 und Studie 2). Allerdings kann dies auch daran gelegen haben, dass das Videomaterial in der männlichen Bedingung eine weniger stereotype „maskuline" Person darstellte und somit als wenig repräsentativ für die Kategorie „Mann" wahrgenommen wurde.

Dass die Kompetenzmanipulation keinen Einfluss auf die Kompetenzbeurteilung am Ende des Vortrags hatte, kann darauf hinweisen, dass die Erfassung der Kompetenzbeurteilung einem subjektiven Maßstab folgt, sodass sich eine Form des *shifting standards* (Biernat et al., 1991; Foschi, 2000) zeigte. Ursprünglich wurde die *shifting standards theory* herangezogen, um die Bewertung von Mitgliedern unterschiedlich stereotypisierter Gruppenmitglieder zu erklären. Beispielsweise wird mit subjektiven Maßstäben (Skala von „sehr klein" bis „sehr groß") die Körpergröße von Männern und Frauen ähnlicher beschrieben als mit objektiven (Schätzung in Zentimetern). Erklärt werden kann dies damit, dass bei subjektiven Maßstäben andere Vergleichsgruppen herangezogen werden, also Frauen werden mit anderen Frauen verglichen und Männer mit anderen Männern. Für die erzielten Studienbefunde heißt dies, dass in der Galatea-Bedingung ein höherer Standard bei der Kompetenzbewertung angelegt wurde als in der Golem-Bedingung. Eine Lehrperson, deren Examen mit „sehr gut" bewertet wurde, wird anhand dieser Note subjektiv bewertet. So können z. B. Gedanken gewesen sein, dass die Lehrperson für „sehr gut" nicht „gut genug" im Vortrag war – analog könnte die Lehrperson mit der Note „befriedigend" als „besser als erwartet" erachtet worden sein, was zu positiveren subjektiven Bewertungen geführt haben könnte. Eine Beurteilung der Kompetenz durch eindeutig objektive Maße zur Überprüfung der Kompetenzmanipulation müsste in ähnlich angelegten Studien künftig berücksichtigt werden.

Weiterhin methodisch kritisch anzumerken ist zum einen die kleine Proband*innen-Zahl ($N = 62$; wegen fehlender Werte 61 für die Regressionsanalysen) und die damit einhergehende geringe statistische Power (vgl. Döring &

Bortz, 2016; Hager, 2004). Dies steht aber in Relation zum sehr aufwendigen Studiendesign und der aufwendigen Auswertung der Videos. Die Auswertung der aufgezeichneten Videos wurde zusätzlich durch die Perspektive der Kameras beeinträchtigt. So waren einige Verhaltensweisen nicht eindeutig zu beobachten (z. B. Zurücklehnen) und die bei Feldman und Prohaska (1979) vorgenommenen Maßeinheiten (z. B. Körperorientierung in Grad) konnten hier nicht verwendet werden. Darüber hinaus unterschieden sich die Kompetenzbedingungen hinsichtlich der Proband*innen-Zusammensetzung: So gab es mehr männliche und mehr ältere Proband*innen in der Galatea-Bedingung, was zu Konfundierungen geführt haben könnte (vgl. u. a. Döring und Bortz, 2016) – wobei das Alter in den multivariaten Analysen als Kontrollvariable berücksichtigt wurde, das Geschlecht aufgrund der sehr geringen männlichen Teilnehmendenzahl ($n = 12$) nicht. Außerdem sind Versuchsleitungseffekte (Klein et al., 2012; Rosenthal, 1966, 2009; Rosenthal & Fode, 1973) nicht auszuschließen, wobei diesen mit einem Leitfaden für das Vorgehen begegnet wurde.

Eine besonders wichtige Einschränkung, die Beachtung finden muss, ist, dass die Cover-Story (Supervision von Lehrkräften) die Proband*innen nur indirekt in eine Lernendenposition brachte. Die Proband*innen hatten nicht die Aufgabe, die Inhalte des Vortrags zu lernen, sondern die Lehrperson zu bewerten. Dies könnte erklären, warum es keine signifikanten Unterschiede im Wissenstest zwischen der Golem- und Galatea-Bedingung gab. Weitergehend ist einschränkend anzumerken, dass aufgrund der fehlenden Testung zum Vorwissen es unklar ist, ob die Proband*innen nicht doch einen höheren Wissenszuwachs in der Galatea-Bedingung hatten. Für das nonverbale Verhalten jedoch sollte sich aufgrund der Cover-Story kein Unterschied ergeben, da Lernende auch immer die Lehrpersonen sowie das Unterrichtsangebot (bewusst oder unbewusst) evaluieren und bewerten (vgl. Nickel, 1985). Um dies jedoch mit Sicherheit aussagen zu können, bedarf es weitergehender Untersuchungen.

Positiv an der Studie sind jedoch folgende Punkte hervorzuheben: Es handelte sich um ein experimentelles Vorgehen, sodass Kausalaussagen möglich sind. Neben der Galatea- und der Golem-Bedingung wäre eine Kontrollgruppe (ohne Erwartungsmanipulation) sinnvoll gewesen, was jedoch eine größere Stichprobe erfordert hätte. Weitergehend wurden keine üblichen Selbsteinschätzungsverfahren genutzt, sondern Verhaltensmaße betrachtet. Verhaltensbeobachtungen haben gegenüber Befragungen eine Reihe von Vorteilen, werden jedoch aufgrund des Aufwands deutlich seltener genutzt (s. u. a. Döring & Bortz, 2016), obwohl sie „in besonderem Maße dem Anspruch der Psychologie gerecht [werden], menschliches Verhalten und Erleben zu beschreiben und zu erklären“ (Weber, 2012, S. 34).

Die bisherigen drei Studien waren Laboruntersuchungen, um Annahmen des transaktionalen Modells von Nickel (1985; s. Kapitel 7) zu Ursachen und Auswirkungen kognitiver Schemata in Kausalanalysen zu betrachten. In der finalen vierten Studie sollen die Überlegungen zur Wirkung von Lernendenkompetenzerwartungen an Lehrpersonen im realen Lehr-Lern-Setting in einer Feldstudie untersucht werden.

16 Studie 4 – Lernendenerwartungen als Einfluss auf die Lehr-Lern-Situation

Die vierte Studie wurde in Lehrveranstaltungen einer Hochschule durchgeführt, an der sowohl Studierende als auch Dozierende teilnahmen. Eine geplante Untersuchung in allgemeinbildenden Schulen, um Ergebnisse im schulischen und universitären Kontext vergleichen zu können, konnte leider nicht realisiert werden. Die Studie wurde aufgrund der geplanten zeitlich parallelen Erhebungen in Schulen und Universität so angelegt, dass sie auch mit Jugendlichen im Schulkontext hätte durchgeführt werden können.

16.1 Hypothesen

Nachdem die dritte Studie einen weiteren Einblick in die Wirkung von Kompetenzerwartungen auf das Verhalten der Lernenden gibt und zusammen mit den Studien von Feldman und Prohaska (1979) sowie Jamieson et al. (1987) geschlussfolgert werden kann, dass sich das Verhalten von Lernenden an ihre Kompetenzerwartungen an Lehrpersonen anpasst, sollen in einem weiteren Schritt Zusammenhänge in einer Feldstudie betrachtet werden. In den bisherigen Befunden konnte gezeigt werden, dass Lernende, die eine Lehrperson für kompetent halten, diese positiver bewerten (Feldman & Prohaska, 1979; Feldman & Thiess, 1982; Lewandowski et al., 2012; Reavis, 1979; Reber et al., 2017; Towler & Dipboye, 2006), den Unterricht interessanter finden (Feldman & Prohaska, 1979; Feldman & Theiss, 1982), sich wohler fühlen (positiver Affekt; Edwards et al., 2009) und motivierter waren, an einer potenziellen Lehrveranstaltung dieser Lehrperson teilzunehmen (Kowai-Bell et al., 2011, 2012). Die Zusammenhänge der

Ergänzende Information Die elektronische Version dieses Kapitels enthält Zusatzmaterial, auf das über folgenden Link zugegriffen werden kann https://doi.org/10.1007/978-3-658-39444-8_16.

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_16

Lernendenkompetenzerwartungen mit dem erwünschten Erleben sollen repliziert werden, sodass folgende Hypothesen aufgestellt werden:

Hypothesenkomplex 1: Lernende, die eine Lehrperson beim ersten Eindruck für kompetenter halten, unterscheiden sich von Lernenden, die eine Lehrperson für weniger kompetent halten, in einer Lehrveranstaltungssitzung beim Erleben von (1.) Freude, (2.) Frustration und Unsicherheit, (3.) Langeweile, (4.) Interesse und (5.) Störungen. Außerdem bewerten sie die Lehrperson hinsichtlich (6.) Kompetenz, (7.) Wärme, (8.) Klassenführung und (9.) Beziehungsgestaltung anders als Lernende, die eine Lehrperson für inkompetent halten.

Zusätzlich zu bisherigen Studien sollen sowohl die Individual- als auch die Gruppenebene berücksichtigt werden (s. Kapitel 11). Die Hypothesen beziehen sich auf beide Ebenen. Da es bei der gemeinsamen Betrachtung der beiden Ebenen zu unterschiedlichen Ergebnissen kommen kann (z. B. positiver Zusammenhang auf Gruppenebene, aber negativer auf Individualebene; vgl. u. a. Eid et al., 2017), sind die Hypothesen ungerichtet.

Die Befunde zu den Wirkungen von Lernendenkompetenzerwartungen auf die Lehrpersonen sind – wie bereits an vielen Stellen erwähnt – sehr rar. So berichteten Lehrpersonen, die sich eine Lehrsituation vorstellten, sich wohler im Umgang mit Lernenden zu fühlen, wenn sie glaubten, die Lernenden bewerteten die Lehrperson positiv (Kowai-Bell et al., 2012). Weitergehend wurden vermeintlich warme Lehrpersonen am Ende eines Vortrags von unabhängigen Beobachter*innen (ohne Erwartungen) positiver bewertet (Herrell, 1971). Besonders hervorzuheben sind allerdings die Ergebnisse von Feldman und Prohaska (1979), die zeigten, dass das von ihnen gefundene prototypische (zugewandtere) Verhalten von Lernenden dazu führte, dass sich die Lehrpersonen fröhlicher ($d = 0.845$), wärmer ($d = 1.171$) und weniger unzufrieden ($d = 0.760$) fühlten. Außerdem beurteilten sie die Lernenden als enthusiastischer ($d = 1.923$) und sympathischer ($d = 1.076$). Beim nonverbalen Verhalten der Lehrpersonen gab es keine signifikanten Unterschiede. Bei Feldman und Theiss (1982) stellte sich jedoch kein Effekt der Lernendenkompetenzerwartungen auf die Lehrpersonen ein. Drei von vier Studien liefern Indizien für einen spezifischen Einfluss, allerdings sind aufgrund der wenigen Befunde die folgenden Zusammenhangshypothesen ungerichtet:

Hypothesenkomplex 2: Lehrpersonen, die im Mittel beim ersten Eindruck für kompetenter gehalten werden, erleben ein anderes Ausmaß an Emotionen, sowohl (1.) positive als auch (2.) negative, (3.) an Flow, (4.) der eigenen Kompetenz und (5.) Wärme. Außerdem unterscheiden sie sich in der Bewertung (6.) der Störungswahrnehmung, (7.) der eigenen Klassenführung und (8.) Beziehungsgestaltung von Lehrpersonen, die im Mittel beim ersten Eindruck für weniger kompetent gehalten werden.

Zur Bewertung des Verhaltens und des Unterrichts werden u. a. Variablen, die im Kontext von Klassenführung im weiten Sinne (vgl. Hannemann, 2020; Helmke, 2017) Verwendung finden, herangezogen. Klassenführung umfasst nach Helmke und Helmke (2014, S. 9) „Konzepte, Strategien und Techniken, die dem Ziel dienen, einen störungsfreien und reibungslosen Unterrichtsverlauf zu ermöglichen und damit aktive Lernzeit zu maximieren: durch Regeln und Prozeduren, die Allgegenwärtigkeit der Lehrkraft, den Aufbau erwünschten Verhaltens und einen angemessenen Umgang mit Störungen“. Klassenführungskompetenz ist ein wichtiger Kompetenzbereich für Lehrpersonen (vgl. u. a. Baumert & Kunter, 2006; Hannemann, 2020; Helmke, 2017; Kunter et al., 2011). Insbesondere der Aspekt der Störungswahrnehmung wird in der durchgeführten Untersuchung fokussiert, weil dies als Marker für die (bewusste und unbewusste) Nutzung des Unterrichtsangebots (vgl. Angebots-Nutzungs-Modelle; Helmke, 2017; Reusser & Pauli, 2010) bewertet werden kann, ohne Lernerfolge heranzuziehen, die auch zu Beginn einer Lehrveranstaltung noch nicht vorliegen (können). So reduzieren Störungen die aktive Lernzeit, die für den Lernerfolg wichtig ist (vgl. Helmke, 2017).

Analog zu den bisherigen Studien sollen die Zusammenhangsanalysen auch mit den Kontrollvariablen Wärmebeurteilung, die der Kompetenzbeurteilung vorausgeht (*stereotype content model*; s. Kapitel 5), und der Attraktivitätsbeurteilung (*„beatuy is good“ stereotype*; Dion et al., 1972; Eagly et al., 1991; Langlois et al., 2000) explorativ betrachtet werden.

16.2 Methodik

16.2.1 Stichprobe

Es wurden 42 Lehrveranstaltungen der Technischen Universität Braunschweig besucht und sowohl Studierende als auch Dozierende befragt. In den 42 untersuchten Veranstaltungen nahmen 41 Dozierende teil. Von den 41 Dozierenden gaben 15 an, männlich zu sein, 25, weiblich zu sein, und eine Person machte keine Angabe zu ihrem Geschlecht. Sie waren im Durchschnitt zum Zeitpunkt der Teilnahme 38.9 Jahre alt ($SD = 11.35$; Spannweite: 26–62; 3 ohne Angabe) und lehrten (teils geschätzt) 16.2 Semester ($SD = 13.7$; 2 ohne Angabe).

Insgesamt liegen 997 Datensätze von Studierenden vor, die ihre Dozierenden und die jeweilige Lehrveranstaltung beurteilten. Proband*innen, die zu einem der zwei Erhebungszeitpunkte (s. u.) keinen Fragebogen ausgefüllt haben, wurden für die Auswertungen nicht berücksichtigt. Dadurch verblieben 871 Datensätze. Durchschnittlich gab es 20.7 studentische Angaben pro Lehrveranstaltung ($SD = 13.8$; Spannweite: 7–78). Mehrfachteilnahmen von Studierenden waren möglich, wenn sie mehr als eine der untersuchten Lehrveranstaltungen besuchten.

Auf Basis aller Datensätze inklusive Mehrfachteilnahmen ($N = 871$): Die Studierenden waren im Mittel 22.7 Jahre alt ($SD = 4.08$; Spannweite: 18–58), wobei 33 keine Angabe machten. Es gaben 221 Studierende an, männlich zu sein (25 %), und 612 der Befragten waren weiblich (70 %). Eine andere Geschlechtsidentität gaben sechs Personen an und 32 machten keine Angabe. Die meisten ($n = 537$, 62 %) studierten primär Fächer der Geistes- und Erziehungswissenschaften.

Werden nur die eindeutigen Ersteilnahmen ($N = 663$; 76 % aller Teilnahmen) berücksichtigt, ergibt sich ein ähnliches Bild: Die Studierenden waren im Mittel 22.6 Jahre alt ($SD = 4.00$; Spannweite: 18–58), wobei 14 keine Angabe machten. Es gaben 161 Studierende an, männlich zu sein (24 %), und 476 der Befragten waren weiblich (72 %). Eine andere Geschlechtsidentität gaben fünf Personen an und 21 machten keine Angabe. Die Mehrheit ($n = 398$; 60 %) studierte primär Fächer der Geistes- und Erziehungswissenschaften. Die Angaben der Erstteilnahmen unterschieden sich bezüglich der Geschlechtsaufteilung (männlich oder weiblich; $\chi^2(1) = 2.191$, $p = .139$, $\varphi = .051 \triangleq r_S$), des Alters ($t(836) = 1.314$, $p = .189$; $d = 0.109$ entspricht $r = .054$) und des Anteils der Studierenden im Bereich der Geistes- und Erziehungswissenschaften ($\chi^2(1) = 7.420$, $p = .006$, $\varphi = -.095 \triangleq r_S$) nicht bedeutsam ($|r| \geq .10$ für einen kleinen Effekt; Cohen, 1988) von den

Angaben, die mindestens ein zweites Mal (bei Mehrfachteilnahme) oder ein unbekanntes Mal (ohne Angabe bisheriger Teilnahmen) gemacht wurden. Zudem ist theoretisch nicht davon auszugehen, dass die Angaben der Proband*innen bei freiwilliger Mehrfachteilnahme die Ergebnisse verzerren. Deswegen werden für die hypothesenprüfenden Analysen alle Datensätze ($N = 871$) herangezogen.

16.2.2 Vorgehen

Die Lehrveranstaltungen wurden zu Beginn des Sommersemesters 2018 sowie Wintersemesters 2018/19 in der maximal fünften Lehrveranstaltungssitzung besucht. Der Beginn einer Lehrveranstaltung wurde gewählt, weil sich dann Kompetenzerwartungen besonders bedeutsam zeigen (vgl. Jussim und Harber, 2005; Raudenbusch, 1984). Zum Start der Sitzung wurden die Studierenden über die Freiwilligkeit, die Anonymität und den Datenschutz aufgeklärt. Zudem wurde beschrieben, die Studie würde sich mit der „Wahrnehmung von Klassenführungsvariablen in der Hochschule“ beschäftigen. Die Studierenden sollten anschließend einen ersten Fragebogen beantworten, in dem retrospektiv der erste Eindruck von der Lehrperson erfasst wurde sowie die Beurteilung der aktuellen Attraktivität der Lehrperson.

Das retrospektive Erfragen ist unter der Bedingung, dass der erste Eindruck recht stabil ist (vgl. Schweer, 2019), eine Methode zur Erfassung der Kompetenzerwartung, die jedoch Schwächen aufweist. So kann die Stabilität der Repräsentation auch auf Erwartungseffekte, also selbsterfüllende Prophezeiungen, und damit auch über das Streben nach kognitiver Konsistenz (vgl. Kruglanski et al., 2018) zurückzuführen sein. Studien zur Akkuratheit retrospektiver Instrumente zur Erfassung interpersonaler Kompetenzerwartungen konnten nicht detektiert werden. Da für Feldstudien in Schulen und Hochschulen die Lernendenkompetenzerwartungen an Lehrpersonen nicht bzw. nur unter sehr kontrollierten und besonderen Bedingungen zeitnah zum Moment des ersten Eindrucks erfasst werden können, mussten sie retrospektiv erhoben werden. Ob dies die tatsächlichen Kompetenzerwartungen des ersten Eindrucks widerspiegelten, bleibt offen, jedoch ist anzunehmen, dass potenzielle kognitive Verzerrungen bei allen Gruppen ähnlich verlaufen. Durchschnittlich lag der erste Eindruck der befragten Studierenden ($N = 871$) von der jeweiligen dozierenden Person 6.37 Monate ($SD = 10.60$) zurück, also zumeist zeitlich nah an der Erhebung. Zusammenfassend kann also angenommen werden, dass es sich bei der hier vorliegenden retrospektiven Erfassung der Lernendenkompetenzerwartungen zumindest um eine Näherung der Erwartungen handelt.

Im Anschluss an den ersten Erhebungsteil für die Studierenden führte die lehrende Person in Abwesenheit der Versuchsleitung die Lehrveranstaltungssitzung regulär durch. Am Ende der Lehrveranstaltungssitzung sollten die Studierenden und Dozierenden die Sitzung und ihr Erleben währenddessen beurteilen. Die Versuchsleitung war zumeist der Autor der vorliegenden Arbeit. Aus organisatorischen Gründen (z. B. aufgrund von zwei sich überschneidenden Untersuchungen) war dies nicht immer der Fall. Die Auswertung der Daten erfolgte mit IBM SPSS 25 und MPlus 8.

16.2.3 Instrumente im Original

Im Folgenden werden die Variablen, die in der Untersuchung verwendet wurden, in der im Fragebogen präsentierten Reihenfolge beschrieben. Die in den Beschreibungen der Instrumente angegebenen internen Konsistenzen beziehen sich auf die zitierten Originalquellen bzw. der Validierungsstudien. Die erzielten Reliabilitäten (Cronbachs Alpha) in der durchgeführten Studie sind für die Studierenden in Tabelle 16.3 und für die Dozierenden in Tabelle 16.7 angegeben. Beispielitems sind in der Tabelle 16.1 für dir Studierenden und in Tabelle 16.2 für die Dozierenden aufgeführt (Fragebögen und Itemanalysen s. Zusatzmaterial).

16.2.3.1 Studierendenfragebogen zu Beginn der Lehrveranstaltungssitzung

Kompetenzeindruck (erster Eindruck): Die Einschätzung der Kompetenz beim ersten Eindruck wurde mittels der vier Adjektive der vorherigen Studien (kompetent, effizient, fähig, intelligent) auf einer siebenstufigen Likert-Skala (1 = „stimme gar nicht zu“ bis 7 = „stimme vollkommen zu“) erfasst.

Wärmeeindruck (erster Eindruck): Die Einschätzung der Wärme beim ersten Eindruck wurde mit vier Adjektiven der vorherigen Studien (warmherzig, gutmütig, wohlwollend, sympathisch) in der gleichen Liste wie die Kompetenzbeurteilung mit einer siebenstufigen Likert-Skala (1 = „stimme gar nicht zu“ bis 7 = „stimme vollkommen zu“) erfasst.

Attraktivitätsbeurteilung: Die Einschätzung der aktuellen physischen Attraktivität wurde ähnlich wie in der ersten und dritten Studie mit folgendem Item auf einer siebenstufigen Likert-Skala (1 = „sehr unattraktiv“ bis 7 = „sehr attraktiv“) erhoben: „Wie optisch attraktiv empfinden Sie die Dozierende/den Dozierenden der Lehrveranstaltung momentan?“

Tabelle 16.1 Verwendete Instrumente und Skalen mit Beispielitems für die Studierenden-Fragebögen zur Bewertung der Veranstaltungssitzung (Studie 4)

Verwendete Instrumente		Original
Skala (Itemanzahl)	Beispielitem	Beispielitem im Original[a] (Itemanzahl)
Studierendenfragebogen		
Kompetenz- und Wärmebeurteilung nach Studien 1–3		
Kompetenz (4)	Kompetent	Kompetent (4)
Wärme (4)	Sympathisch	Sympathisch (4)
Motivationales und emotionales Erleben nach Wegner et al. (2013)		
Freude (4)	Die Seminarsitzung hat mich sehr begeistert.	Der Biologieunterricht der letzten Stunden hat mich sehr begeistert. (4)
Frustration (4, ein Item zu „Langeweile")	Ich wusste in der Seminarsitzung nicht, was Sache ist.	Ich wusste in den letzten Biologiestunden nicht, was Sache ist. (5)
Langeweile (5, ein Item aus „Frustration")	Ich habe bemerkt, dass ich in der Seminarsitzung oft geistig abwesend war.	Ich habe bemerkt, dass ich in den letzten Biologiestunden oft geistig abwesend war. (4)
Interesse (4)	Das Thema der Seminarsitzung hat mich richtig interessiert.	Interessiertheit: Die letzten Biologiestunden haben mich richtig interessiert. (4)
Störungswahrnehmung und Klassenführung nach Wettstein, Ramseier et al. (2016), ergänzt um Störungen in Anlehnung an Keller (2010)		
Störungen (5, 2 Items zu „Teilabsentismus")	Einige Studierende schwatzten mit anderen Studierenden außerhalb von Gruppenarbeiten.	Nicht aggressives Verhalten: Einige Kinder schwatzen, während diese Lehrperson etwas erklärt. (4, inkl. Ergänzungen 14)
Störungsmissmanagement (4)	Die/der Dozierende musste oft eingreifen, um Aufmerksamkeit zu bekommen.	Störungen des methodisch-didaktischen Settings: Diese Lehrperson muss oft eingreifen, um Aufmerksamkeit zu bekommen. (4)
Beziehung (2)	Ich mochte die Dozierende/den Dozierenden.	Ich mag diese Lehrperson. (6)
Klassenführung (0)	Die/der Dozierende hatte Überblick über das, was im Seminar geschah.	Diese Lehrperson hatte Überblick über das, was in der Klasse geschieht. (3)
Teilabsentismus (in Anlehnung an Keller, 2010)		
Teilabsentismus (2, aus „Störungen")	Einige Studierende kamen zu spät.	

Anmerkungen. [a] Bei abweichendem Skalennamen zum Original wurde dieser vor dem Beispielitem, getrennt durch einen Doppelpunkt, benannt.

Tabelle 16.2 Verwendete Instrumente und Skalen mit Beispielitems für die Dozierenden-Fragebögen zur Bewertung der Veranstaltungssitzung (Studie 4)

Verwendete Instrumente		Original
Skala (Itemanzahl)	Beispielitem	Beispielitem im Original[a] (Itemanzahl)
Dozierendenfragebogen		
Affekt nach Rahm et al. (2017)		
Positiver Emotionen (6)	Angenehm	Angenehm (6)
Negativer Emotionen (6)	Unangenehm	Unangenehm (6)
Kompetenz- und Wärmewahrbeurteilung nach Studien 1–3		
Kompetenz DLV (4)	Kompetent	Kompetent (4)
Wärme DLV (4)	Sympathisch	Sympathisch (4)
Flowerleben nach Rheinberg et al. (2003)		
Flowerleben (10)	Ich fühlte mich optimal beansprucht.	Ich fühlte mich optimal beansprucht. (10)
Störungswahrnehmung und Klassenführung nach Wettstein, Ramseier et al. (2016), ergänzt um Störungen in Anlehnung an Keller (2010)		
Störungen (5, 2 Items zu „Teilabsentismus“)	Einige Studierende schwatzten mit anderen Studierenden außerhalb von Gruppenarbeiten.	Nicht aggressives Verhalten: Einige Kinder schwatzen, während diese Lehrperson etwas erklärt. (4, inkl. Ergänzungen 14)
Störungsmissmanagement (4)	Ich muss oft eingreifen, um Aufmerksamkeit zu bekommen.	Störungen des methodisch-didaktischen Settings: Ich musste oft eingreifen, um Aufmerksamkeit zu bekommen. (4)
Beziehung (2)	Ich mag die Studierenden des Seminars.	Ich mag meine Klasse. (6)
Klassenführung (0)	Ich hatte den Überblick über das, was im Seminar geschah.	Diese Lehrperson hatte Überblick über das, was in der Klasse geschieht. (3)
Teilabsentismus (in Anlehnung an Keller, 2010)		
Teilabsentismus (2, aus „Störungen“)	Einige Studierende kamen zu spät.	

Anmerkungen. [a] Bei abweichendem Skalennamen zum Original wurde dieser vor dem Beispielitem, getrennt durch einen Doppelpunkt, benannt.

16.2.3.2 Studierendenfragebogen zum Ende der Lehrveranstaltungssitzung

Kompetenzbeurteilung in Lehrveranstaltungssitzung: Die Einschätzung der Kompetenz innerhalb der Lehrveranstaltungssitzung wurde mittels der vier Adjektive der vorherigen Studien (kompetent, effizient, fähig, intelligent) auf einer siebenstufigen Likert-Skala (1 = „stimme gar nicht zu" bis 7 = „stimme vollkommen zu") erfasst.

Wärmebeurteilung in Lehrveranstaltungssitzung: Die Einschätzung der Wärme innerhalb der Lehrveranstaltungssitzung wurde mit vier Adjektiven der vorherigen Studien (warmherzig, gutmütig, wohlwollend, sympathisch) in der gleichen Liste wie die Kompetenzbeurteilung mit einer siebenstufigen Likert-Skala (1 = „stimme gar nicht zu" bis 7 = „stimme vollkommen zu") erfasst.

Motivationales und emotionales Erleben: Das motivationale und emotionale Erleben der Studierenden während der Lehrveranstaltungssitzung wurde mit einem für die Hochschule adaptierten Fragebogen von Wegner (2009, nach Wegner et al., 2013) erfasst. Der ursprüngliche Fragebogen fokussierte den Biologieunterricht in der Schule und besteht aus vier Subskalen: Freude (4 Items, $\alpha \geq .92$), Frustration (5 Items, $\alpha \geq .75$), Langeweile (4 Items, $\alpha \geq .76$) und Interessiertheit (4 Items, $\alpha \geq .83$). Freude ist dabei bestimmt als das Ausmaß des Erlebens von Spaß, Begeisterung und Freude. Frustriert sein und das Erleben von Unsicherheit gehen in die Skala Frustration ein. Langeweile ist bestimmt als das Gefühl, dass die Zeit nicht vergeht, sowie, keine (sinnvolle) Beschäftigung zu haben, und die Skala Interessiertheit beinhaltet Interesse und Neugier in Bezug auf das Lehrthema. Die Proband*innen sollten auf einer vierstufigen Likert-Skala ihre Zustimmung angeben (1 = „stimmt gar nicht" bis 4 = „stimmt genau").

Störungswahrnehmung und Klassenführung: Die Störungs- und Klassenführungswahrnehmung wurde auf Basis des Fragebogens zur Wahrnehmung von Unterrichtsstörungen aus Lehrkräfte- und Schüler*innenperspektive von Wettstein, Ramseier et al. (2016; vgl. Wettstein, Scherzinger & Wyler, 2016) konstruiert. Der Fragebogen wurde gewählt, da er sowohl eine Schüler*innen- als auch eine Lehrkräfte-Version bereithält. Es wurden die Subskalen „Nicht aggressives Verhalten der Schülerinnen und Schüler" (Schüler*innen: $\alpha = .88$, Lehrkräfte: $\alpha = .72$) mit einer vierstufigen Likert-Skala zur Einschätzung der Häufigkeit (1 = „nie" bis 4 „sehr oft"), „Störungen des methodisch-didaktischen Settings" (Schüler*innen: $\alpha = .87$, Lehrkräfte: $\alpha = .82$), „Beziehung" (Schüler*innen: $\alpha = .95$, Lehrkräfte: $\alpha = .55$) und „Klassenführung" (Schüler*innen: $\alpha = .82$, Lehrkräfte: $\alpha = .58$) mit einer vierstufigen Likert-Skala zur Zustimmung (1 = „stimmt nicht" bis 4 = „stimmt") herangezogen. Die Items wurden so

ausgewählt, dass sie sowohl für den schulischen als auch für den Hochschulkontext augenscheinlich sinnvoll sind. Die Items wurden für den Hochschulkontext umformuliert sowie um ausgewähltes, nicht aggressives Störungsverhalten, das in Keller (2010) formuliert ist, ergänzt. Außerdem wurde die Option „nicht sinnvoll beantwortbar" hinzugefügt.

Soziodemografische und weitere Variablen: Es wurden das Geschlecht erfragt, das Alter, das Fachsemester[1], die Fakultät(en) des Studiengangs, die Häufigkeit des Besuchs einer Lehrveranstaltung vor dem Sommersemester 2018[2] bei der entsprechenden dozierenden Person, die Zeit, wie lang der erste Eindruck zurückliegt, und wie häufig bereits an der Studie teilgenommen wurde.

16.2.3.3 Dozierendenfragebogen (zum Ende der Lehrveranstaltungssitzung)

Positive und negative Emotionen: Das Erleben von positiven und negativen Emotionen wurde mit der deutschen Version der *scale of positive and negative experience* (SPANE) von Rahm et al. (2017) erhoben. Beim SPANE werden abwechselnd positive (z. B. „gut"; $\alpha = .87$) und negative Beschreibungen (z. B. „unangenehm"; $\alpha = .82$) präsentiert. Da nur das Erleben innerhalb der Lehrveranstaltungssitzung erfasst werden sollte, wurde die Intensität mit einer fünfstufigen Likert-Skala (1 = „gar nicht" bis 5 = „außerordentlich") – analog zu Friedrich (2017) – erhoben.

Beurteilung der eigenen Kompetenz in der Lehrveranstaltungssitzung: Die Selbsteinschätzung der Kompetenz innerhalb der Lehrveranstaltungssitzung wurde mittels der vier Adjektive der vorherigen Studien (kompetent, effizient, fähig, intelligent) auf einer siebenstufigen Likert-Skala (1 = „stimme gar nicht zu" bis 7 = „stimme vollkommen zu") erfasst. Die Instruktion wurde für eine Selbsteinschätzung angepasst.

Beurteilung der eigenen Wärme in der Lehrveranstaltungssitzung: Die Selbsteinschätzung der Wärme innerhalb der Lehrveranstaltungssitzung wurde mit vier Adjektiven der vorherigen Studien (warmherzig, gutmütig, wohlwollend, sympathisch) in der gleichen Liste wie die Kompetenzbeurteilung mit einer siebenstufigen Likert-Skala (1 = „stimme gar nicht zu" bis 7 = „stimme

[1] Da einige Studierende ausschließlich das Fachsemester des Masterstudiums und andere das Fachsemester des Masterstudiums addiert um die Fachsemester des Bachelorstudiums berichteten und somit die Angaben nicht vergleichbar sind, wurde diese Variable für die Beschreibung der Stichprobe nicht berücksichtigt.

[2] Da die Frage nicht für den Durchgang im Wintersemester 2018/19 angepasst wurde, wurde sie in den Analysen nicht berücksichtigt.

vollkommen zu“) erfasst. Die Instruktion wurde für eine Selbsteinschätzung angepasst.

Flowerleben: Das Flowerleben wurde mit der 10-Items-Flow-Kurzskala (FKS) von Rheinberg et al. (2003, 2019) erhoben. Die Skala besteht aus den Subskalen „glatter automatischer Verlauf“ ($\alpha = .92$) und „Absorbiertheit“ ($\alpha = .80$), die beide eine siebenstufige Zustimmungs-Likert-Skala nutzen (1 = „trifft nicht zu“ bis 7 = „trifft zu“). Rheinberg et al. (2003) argumentieren auch für einen Generalfaktor ($\alpha = .90$), der in den Analysen dieser Arbeit genutzt wurde.

Störungswahrnehmung und Klassenführung: Die Störungswahrnehmung und die Wahrnehmung der eigenen Klassenführung wurde mit einer für Dozierende parallel gestalteten Version der Studierendenfragebögen in Anlehnung an Wettstein, Ramseier et al. (2016; vgl. Wettstein, Scherzinger & Wyler, 2016) erfasst (s. o.).

Soziodemografische und weitere Variablen: Es wurden das Geschlecht, das Alter, die Lehrerfahrung in Semestern, die zugeordnete Fakultät der Lehrveranstaltung und die Anzahl der bisherigen Sitzungen der Lehrveranstaltung erhoben.

16.2.3.4 Faktorenanalysen und finale Instrumente

Da alle Instrumente für die Studierenden ursprünglich für den schulischen Kontext konstruiert wurden[3], wurden die Items und Skalen anhand der studentischen Stichprobe überprüft. Die Basis für die Item- und Faktorenanalysen bildeten die eindeutigen Ersteilnahmen ($N = 663$), wobei die finale Faktorenstruktur für alle studentischen Angaben (inkl. Mehrfachteilnahmen; $N = 871$) überprüft wurde.

Zuerst wurden die Items auf Basis der Schwierigkeiten ($5 \leq P_i \leq 95$; vgl. Kelava & Moosbrugger, 2020)[4] ausgewählt. Items, bei denen die Option „Nicht sinnvoll beantwortbar“ mit mindestens einem Viertel (25 %) der Antworten entsprach, wurden weitergehend aus den weiteren Analysen ausgeschlossen; diese scheinen nicht relevant für die Hochschule zu sein. Da die Angaben der Studierenden innerhalb von Gruppen strukturiert sind, sollten dann Mehrebenen-Faktorenanalysen berechnet werden ($.054 \leq ICC \leq .423$, $M = .194$, $SD = .074$; s. Zusatzmaterial). Bei Mehrebenen-Faktorenanalysen wird die

[3] Die Auswahl von Items, die für den schulischen Kontext gedacht sind, sollte einen Vergleich zwischen Schule und Hochschule ermöglichen.

[4] Da bei den freiwillig teilnehmenden Dozierenden ein Selektionseffekt vorliegen kann, sodass insbesondere die kompetenten und selbstbewussten Dozierenden sich bereit erklärten (s. a. Tollefson & Wigington, 1986), wurde sich für eine weite Spannweite der Itemschwierigkeit entschieden.

Gesamtvarianz von Beobachtungen zerlegt in die Varianz innerhalb (Individualebene; *within*) und zwischen Gruppen (bzw. Kontexteinheiten; Gruppenebene; *between*), wobei unterschiedliche Faktorenstrukturen zwischen Gruppen- und Individualebene[5] entstehen können (Kleinke et al., 2017). Simulationsstudien zeigten jedoch, dass 50 Level-2-Einheiten (hier: 42 Lehrveranstaltungsgruppen) für den vorliegenden Datensatz (z. B. häufig keine Normalverteilungen) zu gering sind (Holtmann et al., 2016; Hox & Maas, 2001; Hox et al., 2010). Deswegen wurden die Faktorenanalysen ohne Berücksichtigung der hierarchischen Datenstruktur durchgeführt.

Für die Faktorenanalysen wurde ein robustes Weighted-Least-Squares-Verfahren (WLSMV) genutzt, das robust gegen Verletzungen der Normalverteilung und auch für ordinale Daten angemessen ist (Holtmann et al., 2016; Li, 2016). Auch die Nutzung der teilweise wenigen Abstufungen der Likert-Skalen (Bühner, 2010) war ein Grund für das Verfahren für ordinale Daten. Darüber hinaus ist das WLSMV-Verfahren gut für kleine Stichprobengrößen geeignet (Eid et al., 2017). Um die besondere Datenstruktur zu berücksichtigen, wurde außerdem „COMPLEX" als Analysetyp in MPlus herangezogen. „COMPLEX computes standard errors and a chi-square test of model fit taking into account stratification, non-independence of observations, and/or unequal probability of selection" (Muthén & Muthén, 2017, S. 659). Insgesamt wurden also sehr robuste Verfahren für die Bestimmung der Faktorenstruktur herangezogen. Zwei Faktorenanalysen wurden berechnet: eine für den ersten Eindruck (Fragebogen am Anfang der Veranstaltungssitzung) und eine zweite für die Beurteilung der Veranstaltungssitzung (Fragebogen am Ende der Veranstaltungssitzung).

Die konfirmatorische Faktorenanalyse zum ersten Eindruck bestätigte die zweifaktorielle Lösung mit der Wärme- und Kompetenzbeurteilung über die Angaben der Erstteilnahmen ($N = 663$; $\chi^2/df = 2.794$; CFI = .998; TLI = .997; RMSEA = .052; WRMR = 0.560). Über alle Angaben ($N = 871$) ergaben sich auch gute Fitmaße ($\chi^2/df = 3.739$; CFI = .997; TLI = .996; RMSEA =.056; WRMR = 0.644), wenngleich das Verhältnis von χ^2 zu den Freiheitsgraden über dem empfohlenen Cutoff-Wert von 3 liegt (vgl. Weiber & Mühlhaus, 2014).

Für die Maße spezifisch für die Veranstaltungssitzung wurde ein gemeinsames Modell herangezogen. Dies erzielte gute Fitmaße ($\chi^2/df = 1.398$;

[5] Im Folgenden werden die Begriffe „Individualebene" und „Gruppenebene" für hierarchische Analysen genutzt. Für Analysen ohne Berücksichtigung der hierarchischen Datenstruktur werden die Begriffe „individuelle Ebene" und „aggregierte Gruppenwerte" genutzt.

CFI = .981; TLI = .979; RMSEA = .024), jedoch lag der WRMR mit 1.015 zu hoch (vgl. Cook et al., 2009; Yu, 2002). Auf Basis verschiedener explorativer Faktorenanalysen (WLSMV und COMPLEX) der einzelnen Fragebögen wurden folgende Änderungen im Gesamtmodell vorgenommen: Das Item „In der Seminarsitzung war ich eifrig bei der Sache“ wurde dem Faktor „Langeweile“ zugeordnet (ursprünglich „Frustration“). Die Items „Einige Studierende kamen zu spät“ und „Einige Studierende gingen zu früh“ wurden zu einem separaten Faktor „Teilabsentismus“ (vorher im Faktor „Störungen“). Die Items zur Klassenführung wurden aufgrund uneinheitlicher Faktorladungen entfernt („Die/der Dozierende hatte den Überblick über das, was im Seminar geschah“, „Die/der Dozierende merkte, wenn Studierende nicht bei der Sache waren“). Die finale Lösung mit folgenden Faktoren erzielte gute bis sehr gute Fitmaße (χ^2/df = 1.386; CFI = .984; TLI = .982; RMSEA = .024; WRMR = 0.942; mit Attraktivitätsbeurteilung als einzelnes Item: χ^2/df = 1.348; CFI = .984; TLI = .982; RMSEA = .023; WRMR = 0.890): Kompetenzbeurteilung für die Lehrveranstaltungssitzung, Wärmebeurteilung für die Lehrveranstaltungssitzung, Freude, Frustration, Langeweile, Interesse (Bezeichnung im Original: „Interessiertheit“), Störungen (Bezeichnung im Original: „Nicht aggressives Verhalten der Schülerinnen und Schüler“), Teilabsentismus, Störungsmissmanagement (Bezeichnung im Original: „Störungen des methodisch-didaktischen Settings“) und Beziehung (sowie Attraktivitätsbeurteilung; Beispielitems der Skalen s. Tabelle 16.1; alle Items s. Zusatzmaterial). Eine anschließende konfirmatorische Faktorenanalyse über alle Angaben (N = 871) bestätigte die faktorielle Validität (χ^2/df = 1.439; CFI = .983; TLI = .981; RMSEA = .022; WRMR = 0.974; mit Attraktivitätsbeurteilung als einzelnes Item: χ^2/df = 1.397; CFI = .984; TLI = .981; RMSEA = .021; WRMR = 0.921).

16.3 Ergebnisse

16.3.1 Studierende

Die zentralen Kennwerte der Skalen können der Tabelle 16.3 entnommen werden und die (Rang-)Korrelationen der Tabelle 16.4 (Pearson-Korrelationen auf individueller Ebene s. Zusatzmaterial). Die Ergebnisse der Studierenden wurden sowohl auf Individual- als auch auf Gruppenebene betrachtet, um die Mehrebenen-Struktur der Daten zu berücksichtigen: Zur Überprüfung des Hypothesenkomplexes 1 wurden hierarchische Regressionsanalysen auf Basis aller studentischen Angaben (N = 871) berechnet.

Tabelle 16.3 Kennwerte der Studierenden-Variablen (Studie 4)

	N	*M*	*SD*	α	SW	Itemanzahl
*Proband*innen*						
Geschlecht[a]	833	1.73	0.44		.55**	1
Alter	838	22.65	4.08		.71**	1
Erster Eindruck und Attraktivitätsbeurteilung						
Wie lang liegt der erste Eindruck zurück? (in Monaten)	770	6.37	10.58		.64**	1
Kompetenzeindruck[b]	865	5.95	1.11	.92	.81**	4
Wärmeeindruck[b]	861	5.66	1.33	.95	.87**	4
Attraktivitätsbeurteilung[b]	847	4.37	1.20		.93**	1
Beurteilung innerhalb der Lehrveranstaltungssitzung						
Kompetenzbeurteilung in LV[b]	868	6.15	1.04	.92	.76**	4
Wärmebeurteilung in LV[b]	863	5.86	1.26	.96	.83**	4
Freude[c]	866	2.84	0.71	.89	.96**	4
Frustration[c]	866	1.45	0.60	.81	.77**	5
Langeweile[c]	561	1.66	0.43	.84	.91**	4
Interesse[c]	864	2.69	0.70	.85	.98**	4
Störungen[c]	521	1.74	0.47	.70	.94**	5
Teilabsentismus[c]	792	1.67	0.59	.54	.88**	2
Störungsmissmanagement[c]	686	1.31	0.46	.80	.71**	4
Beziehung[c]	659	3.51	0.63	.85	.76**	2

Anmerkungen. Gesamtstichprobe: $N = 871$; Mehrfachteilnahmen waren möglich. α = Cronbachs Alpha; SW = Shapiro-Wilk-Test auf Normalverteilung; LV = Lehrveranstaltungssitzung. [a] Codierung: 1 = männlich, 2 = weiblich. [b] Mögliche Werte: 1–7. [c] Mögliche Werte: 1–4. * $p < .05$. ** $p < .01$.

In der ersten Regressionsanalyse wurde ausschließlich die Kompetenzbeurteilung beim ersten Eindruck (Kompetenzeindruck) als Regressor berücksichtigt. Die Ergebnisse sind in Tabelle 16.5 zu finden. Die Varianz auf der Individualebene konnte durch den Kompetenzeindruck weniger gut aufgeklärt werden ($.000 \leq R^2 \leq .527$) als die auf Gruppenebene ($.052 \leq R^2 \leq .851$). Bedeutsam ($|\beta| \geq .300$) hing der Kompetenzeindruck auf Individualebene mit der Bewertung der Kompetenz ($\beta = .527$, $p < .001$) und Wärme ($\beta = .321$, $p < .001$) der Dozierenden in der Lehrveranstaltungssitzung zusammen. Auf der Gruppenebene ergaben sich bedeutsame Zusammenhänge für alle Regressanden mit Ausnahme der Wahrnehmung von Teilabsentismus ($\beta = -.229$,

Tabelle 16.4 Korrelationen der Studierenden-Variablen auf individueller Ebene und der aggregierten Gruppenwerte (Studie 4)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Geschl.[a]	–	−.09**	−.02	.07	−.07*	.01	.06	−.05	−.05	.05	−.01	−.07*	−.07	−.06	.04	−.02
2	Alter	.01	–	−.03	.13**	.02	.10**	.17**	.07*	.18**	−.12**	−.18**	.20**	−.12**	.04	−.09*	.20**
3	1. Eindr.[b]	.00	−.07	–	.12**	.10**	−.02	.12**	.10**	−.01	−.04	.04	−.02	.09	−.02	.06	.05
4	Komp.[c]	.19	.13	.29	–	.55**	.22**	.76**	.52**	.37**	−.32**	−.30**	.24**	−.30**	−.07	−.30**	.39**
5	Wärme[c]	−.29	.04	.20	.40**	–	.25**	.46**	.82**	.36**	−.34**	−.20**	.18**	−.20**	−.04	−.17**	.48**
6	Attrakt.	−.01	.15	−.19	.18	.30	–	.24**	.24**	.22**	−.20**	.23**	.20**	−.15**	−.11**	−.08*	.26**
7	Komp. LV[d]	.32*	.22	.39*	.87**	.38*	.19	–	.60**	.47**	.40**	−.38**	.34**	−.28**	−.11**	−.33**	.43**
8	Wärme LV[d]	−.11	.10	.29	.57**	.87**	.24	.65**	–	.47**	−.41**	−.29**	.29**	−.24**	−.07	−.23**	.57**
9	Freude	−.02	.27	.07	.61**	.54**	.21	.67**	.67**	–	−60**	−.67**	.68**	−.42**	−.18**	−.30**	.46**
10	Frustration	.06	−.33*	−.10	− .46**	−.46**	−.33*	− .48**	−.58**	− .70**	–	.58**	−.50**	.28**	.12**	.30**	−.41**
11	Langew.	−.19	−.24	−.01	− .52**	−.22	−.33*	−.57**	−.33*	−.79**	.67**	–	−.59**	.49**	.19**	.34**	−.36**
12	Interesse	−.03	.33*	.08	.52*	.40**	.21	.59**	.49**	.87**	−.61**	−.74**	–	−.33**	−.16**	−.22**	.35**
13	Störungen	−.20	.05	−.06	− .46**	−.24	−.10	−.51**	−.32*	−.70**	.41**	.73**	−.53**	–	.34**	.47**	−.33**
14	Teilabs.	−.10	.11	−.10	− .17	−.15	−.04	−.24	−.25	−.30	.25	.37*	−.17	.47**	–	.30**	−.08*
15	SMM	−.32*	−.12	−.09	− .48**	−.23	.07	−.57**	−.44**	−.52**	.36*	.49**	−.34*	.70**	.51**	–	−.28**
16	Beziehung	−.17	.27	.19	.51**	.73**	.29	−.55**	.81**	.65**	−.64**	−.42**	.50**	−.42**	−.22	−.47**	–
17	Gruppengr.[e]	.16	−.23	−.08	−.29	−.26	−.17	−.23	−.19	−.34*	.27	.36*	−.37*	.52**	.07	.36*	−.36*

Anmerkungen. Spearman-Korrelationen: über der Diagonalen auf individueller Ebene ($434 \leq N \leq 865$); unter der Diagonalen auf aggregierter Gruppenebene ($N = 42$). Attrakt. = Attraktivitätsbeurteilung; Langew. = Langeweile; Teilabs. = Teilabsentismus; SMM = Störungsmissmanagement. [a] Geschlecht, Codierung: 1 = männlich, 2 = weiblich. [b] Erster Eindruck: Zurückliegen des ersten Eindrucks in Monaten. [c] Kompetenz- (Komp.) und Wärmeeindruck (Wärme): subjektive Beurteilung beim ersten Eindruck. [d] Kompetenz- (Komp. LV) und Wärmebeurteilung (Wärme LV): subjektive Beurteilung während der Lehrveranstaltungssitzung. [e] Gruppengröße: Anzahl der gültigen teilnehmenden Studierenden. * $p < .05$. ** $p < .01$.

p = .116). Mittelgroße bis große Zusammenhänge (|β| ≥ .500) zeigten sich bei der Kompetenz- (β = .923, p < .001) und Wärmebeurteilung (β = .543, p < .001) der Dozierenden in der Lehrveranstaltungssitzung, der berichteten Freude (β = .604, p < .001) und Langeweile (β = −.689, p < .001), dem Interesse am Lehrveranstaltungsthema (β = .513, p < .001), der wahrgenommenen Beziehung (β = .569, p < .001) sowie der Wahrnehmung von Störungen (β = −.771, p < .001) und des Störungsmissmanagements (β = −.855, p < .001).

In der zweiten Regressionsanalyse gingen die Kompetenz- und Wärmebeurteilung (Kontrollvariable basierend auf dem *stereotype content model*) des ersten Eindrucks (Kompetenz- und Wärmeeindruck) sowie die Attraktivitätsbeurteilung (Kontrollvariable aufgrund des *„beauty is good" stereotype*) als Regressoren ein. Die Ergebnisse der hierarchischen Regressionsanalysen sind in Tabelle 16.6 dargestellt. Insgesamt konnten die Regressionsanalysen viel bis sehr viel Varianz der Regressanden signifikant auf Gruppenebene aufklären (.345 ≤ R^2 ≤ .972); ausgenommen hiervon war der Teilabsentismus (R^2 = .044, p = .443). Auf der Individualebene konnte weniger Varianz aufgeklärt werden (.004 ≤ R^2 ≤ .507). Unter Berücksichtigung der Kontrollvariablen verringerten sich die Zusammenhänge des Kompetenzeindrucks mit einigen Regressanden, was insbesondere auf die Zusammenhänge mit dem Wärmeeindruck zurückzuführen ist. Die Attraktivitätsbeurteilung ist kaum bedeutsam. Auf der Individualebene hing der Kompetenzeindruck nur noch bedeutsam (|β| ≥ .300) mit der Kompetenzbeurteilung der Dozierenden in der Lehrveranstaltungssitzung zusammen (β = .681, p < .001). Während der Kompetenzeindruck in der ersten hierarchischen Regressionsanalyse ohne Kontrollvariablen mittelgroße bis große Zusammenhänge (|β| ≥ .500) mit acht von zehn Regressanden auf der Gruppenebene hatte, ergaben sich in der zweiten Regressionsanalyse mit Kontrollvariablen nur noch mittelgroße bis große Zusammenhänge mit vier Regressanden: Kompetenzbeurteilung der Dozierenden in der Lehrveranstaltungssitzung (β = .858, p < .001), Langeweile (β = −.617, p < .001), Wahrnehmung von Störungen (β = −.764, p < .001) und von Störungsmissmanagement (β = −.797, p < .001). Für die Kompetenzerwartungen – unter Berücksichtigung der Kontrollvariablen – können die Hypothesen 1, 3, 4, 5, 6 und 8 des ersten Hypothesenkomplexes auf Gruppenebene bei mindestens mittelgroßer Effektstärke (|β| > .300) angenommen werden sowie die Hypothese 6 für die Individualebene (s. Tabelle 16.6).

Tabelle 16.5 Ergebnisse der hierarchischen Regressionsanalysen ohne Kontrollvariablen für das Studierenden-Erleben (Studie 4)

			Individualebene (*within*)		Gruppenebene (*between*)	
	N	Konstante *B* (p)	Kompetenz[a] β (*p*)	R^2 (*p*)	Kompetenz[a] β (*p*)	R^2 (*p*)
Kompetenz LV[b]	861	2.186 (.195)	.726 (< .001)	.527 (< .001)	.923 (< .001)	.851 (< .001)
Wärme LV[b]	859	1.992 (.344)	.566 (< .001)	.321 (< .001)	.543 (< .001)	.295 (.007)
Freude	861	−0.061 (.971)	.270 (< .001)	.073 (.016)	.604 (< .001)	.365 (.001)
Frustration	861	11.822 (< .001)	−.214 (< .001)	.046 (.048)	−.471 (< .001)	.222 (.073)
Langeweile	856	15.859 (< .001)	−.191 (< .001)	.037 (.050)	−.689 (< .001)	.474 (.001)
Interesse	858	1.058 (.569)	.143 (.007)	.021 (.180)	.513 (< .001)	.263 (.027)
Störungen	518	16.355 (< .001)	−.166 (< .001)	.028 (.051)	−.771 (< .001)	.595 (.002)
Teilabsentismus	787	7.738 (< .001)	−.019 (.643)	.000 (.817)	−.229 (.116)	.052 (.432)
Störungsmissmanagement	681	20.965 (< .001)	−.153 (< .001)	.023 (.059)	−.855 (< .001)	.731 (< .001)
Beziehung	655	3.267 (.197)	.301 (< .001)	.091 (.008)	.569 (< .001)	.323 (.005)

Anmerkungen. Robuste Maximum Likelihood-Schätzung (MLR). Gesamtstichprobe: $N = 871$; Mehrfachteilnahmen waren möglich. [a] Subjektive Beurteilung beim ersten Eindruck. [b] Subjektive Beurteilung während der Lehrveranstaltungssitzung.

Tabelle 16.6 Ergebnisse der hierarchischen Regressionsanalysen mit Kontrollvariablen für das Studierenden-Erleben (Studie 4)

			Individualebene (*within*)				Gruppenebene (*between*)			
	N	Konstante *B* (*p*)	Kompetenz[a] β (*p*)	Wärme[a] β (*p*)	Attraktivität[b] β (*p*)	R^2 (*p*)	Kompetenz[a] β (*p*)	Wärme[a] β (*p*)	Attraktivität[b] β (*p*)	R^2 (*p*)
Kompetenz LV[c]	832	2.282 (.171)	.681 (< .001)	.014 (.782)	.064 (.011)	.497 (< .001)	.858 (< .001)	.172 (.136)	−.034 (.770)	.887 (< .001)
Wärme LV[c]	829	−0.970 (.448)	.136 (.011)	.599 (< .001)	.068 (< .001)	.507 (< .001)	.115 (< .001)	.926 (< .001)	−.057 (.452)	.934 (< .001)
Freude	830	−2.458 (.145)	.169 (.012)	.109 (.038)	.188 (< .001)	.121 (.001)	.399 (.010)	.517 (.001)	.015 (.898)	.623 (< .001)
Frustration	830	15.020 (< .001)	−.113 (.114)	−.136 (.006)	−.096 (.014)	.071 (.003)	−.147 (.401)	−.545 (< .001)	−.329 (.017)	.638 (< .001)
Langeweile	825	16.561 (< .001)	−.146 (.010)	−.011 (.826)	−.208 (< .001)	.079 (.001)	(< .001) (< .001)	−.031 (.820)	−.245 (.125)	.545 (< .001)
Interesse	827	−0.599 (.761)	.106 (.083)	.002 (.976)	.186 (< .001)	.053 (.003)	.413 (.023)	.238 (.180)	.109 (.464)	.375 (.003)
Störungen	503	19.434 (< .001)	−.191 (.003)	.060 (.359)	−.082 (.140)	.033 (.025)	−.764 (< .001)	−.299 (.011)	−.033 (.824)	.806 (< .001)
Teilabsentismus	761	7.577 (< .001)	−.052 (.408)	.054 (.324)	−.052 (.277)	.004 (.487)	−.152 (.260)	−.059 (.741)	−.059 (.752)	.044 (.443)
Störungsmissmanagement	664	21.606 (< .001)	−.225 (< .001)	.111 (.047)	−.067 (.214)	.036 (.014)	−.797 (< .001)	−.236 (.103)	.058 (.675)	.805 (< .001)
Beziehung	631	−0.741 (.534)	.095 (.174)	.302 (< .001)	.213 (< .001)	.224 (< .001)	.259 (.008)	.890 (< .001)	−.096 (.320)	.972 (< .001)

Anmerkungen. Robuste Maximum Likelihood-Schätzung (MLR). Gesamtstichprobe: $N = 871$; Mehrfachteilnahmen waren möglich. [a] Subjektive Beurteilung beim ersten Eindruck. [b] Attraktivitätsbeurteilung. [c] Subjektive Beurteilung während der Lehrveranstaltungssitzung.

Der Wärmeeindruck hing auf Individualebene ($|\beta| \geq .300$) mit der Beurteilung der Wärme in der Lehrveranstaltungssitzung ($\beta = .599$, $p < .001$) sowie der Beschreibung der Beziehung ($\beta = .302$, $p < .001$) zusammen. Auf Gruppenebene hing der Wärmeeindruck ($|\beta| \geq .500$) hingegen mit der Wärmebeurteilung in der Lehrveranstaltungssitzung ($\beta = .926$, $p < .001$), der Freude ($\beta = .517$, $p = .001$), der Frustration ($\beta = -.545$, $p < .001$) und der Beschreibung der Beziehung ($\beta = .890$, $p < .001$) in der Lehrveranstaltungssitzung zusammen.

16.3.2 Dozierende

In Tabelle 16.7 sind die zentralen Kennwerte der Variablen und die Korrelationen in Tabelle 16.8 gelistet. In Tabelle 16.9 finden sich die Korrelationen mit dem mittleren ersten Eindruck der Studierenden (Kompetenz- und Wärmeeindruck) sowie deren Attraktivitätsbeurteilung der Dozierenden (aggregierte Gruppenwerte).

Tabelle 16.7 Kennwerte der Dozierenden-Variablen (Studie 4)

	N	*M*	*SD*	α	SW	Itemanzahl
Geschlecht[a]	40	1.63	0.49		.62**	1
Alter	38	38.89	11.35		.82**	1
Lehrerfahrung (in Semestern)	39	16.21	13.73		.79**	1
Positive Emotionen	40	3.78	0.60	.83	.94	6
Negative Emotionen	37	1.27	0.36	.81	.79**	6
Kompetenz DLV[b]	41	5.43	0.90	.77	.94	4
Wärme DLV[b]	40	5.36	0.86	.85	.94	4
Flowerleben	40	5.39	0.72	.73	.92	10
Störungen	29	1.50	0.36	.65	.92	5
Teilabsentismus	40	1.66	0.54	.56	.85*	2
Störungsmissmanagement	32	1.45	0.48	.74	.89	4
Beziehung	33	3.77	0.33	.54	.73**	2

Anmerkungen. α = Cronbachs Alpha; SW: Shapiro-Wilk-Test auf Normalverteilung. [a] Codierung: 1 = männlich, 2 = weiblich. [b] Subjektive Beurteilung der Dozierenden von sich selbst während der Lehrveranstaltungssitzung. * $p < .05$. ** $p < .01$.

Tabelle 16.8 Korrelationen der Dozierenden-Variablen (Studie 4)

		2	3	4	5	6	7	8	9	10	11	12	13
1	Geschlecht[a]	−.40*	−.18	−.06	.17	−.13	.27	−.07	.17	.04	.19	.05	.36*
2	Alter		.83**	−.01	−.22	.19	−.12	.14	−.34	−.02	−.29	−.03	−.31
3	Lehrerfahrung (in Semestern)			.11	−.18	.31	−.03	.24	−.30	−.19	.35	.00	.30
4	Positive Emotionen				−.55**	.62**	.30	.64**	−.09	-.28	−.27	.15	−.08
5	Negative Emotionen					−.65**	−.25	−.57**	.47*	.29	.27	.18	−.04
6	Kompetenz DLV[b]						.20	.64**	−.24	−.15	−.51**	−.05	.08
7	Wärme DLV[b]							.31	−.39*	−.29	−.18	.16	.09
8	Flowerleben								−.50**	−.28	−.14	−.02	−.03
9	Störungen									.48**	.15	−.25	.28
10	Teilabsentismus										.17	−.17	.03
11	Störungsmissmanagement											.13	−.04
12	Beziehung												.02
13	Gruppengröße[c]												

*Anmerkungen.*Korrelationen nach Spearman. $23 \leq N \leq 41$. [a] Codierung: 1 = männlich, 2 = weiblich. [b] Subjektive Beurteilung der Dozierenden von sich selbst während der Lehrveranstaltungssitzung. [c] Gruppengröße: Anzahl der gültigen teilnehmenden Studierenden.
* $p < .05$. ** $p < .01$.

Tabelle 16.9 Korrelationen des Dozierenden-Erlebens mit dem studentischen ersten Eindruck und der Attraktivitätsbeurteilung (Studie 4)

	Erster Eindruck und Attraktivitätsbeurteilung der Studierenden[a]			
	Kompetenz			
	r_S	p	Wärme	Attraktivität
Geschlecht[b]	−.291	.069	−.17	.22
Alter	−.040	.811	−.06	−.60**
Lehrerfahrung (in Semestern)	.059	.722	−.05	−.49**
Positive Emotionen	.327*	.039	.09	.24
Negative Emotionen	−.078	.645	.16	−.07
Kompetenz DLV[c]	.318*	.043	.01	.13
Wärme DLV[c]	.205	.203	.48**	.19
Flowerleben	.220	.173	.11	.17
Störungen	−.084	.666	.07	.05
Teilabsentismus	−.060	.714	−.01	−.17
Störungsmissmanagement	−.388*	.028	.12	.22
Beziehung	.223	.212	.16	.09

Anmerkungen. Korrelationen nach Spearman. $29 \leq N \leq 42$. [a] Subjektive mittlere studentische Beurteilung der Dozierenden (aggregierte Gruppenwerte). [b] Codierung: 1 = männlich, 2 = weiblich. [c] Subjektive Beurteilung der Dozierenden von sich selbst während der Lehrveranstaltungssitzung. * $p < .05$. ** $p < .01$.

Die Hypothesen, inwieweit der erste Eindruck der Kompetenz, den Lernende von ihren Dozierenden hatten, mit der Wahrnehmung einer Lehrveranstaltungssitzung der Dozierenden und der Wahrnehmung von sich selbst zusammenhängt, wurden in bivariaten Analysen mit Rangkorrelationen nach Spearman überprüft (s. Tabelle 16.9). Der mittlere erste Kompetenzeindruck hing signifikant mit der Einschätzung der eigenen Kompetenz der Dozierenden in der Lehrveranstaltungssitzung zusammen ($r_S = .318$, $p = .043$; $1-\beta = .527$), außerdem mit dem Erleben positiver Emotionen ($r_S = .327$, $p = .039$; $1-\beta = .551$) und von Störungsmissmanagement ($r_S = -.388$, $p = .028$; $1-\beta = .713$). Weitergehend zeigten die Vorzeichen der Korrelationen, dass der erste Kompetenzeindruck der Studierenden mit einem positiven Erleben der Dozierenden sowie einer besseren Beurteilung einer Lehrveranstaltungssitzung zusammenhing.

Das multivariate Zusammenspiel vom ersten Kompetenz- und Wärmeeindruck (Kontrollvariable basierend auf dem *stereotype content model*), den Studierende von den Dozierenden hatten, sowie der Attraktivität (Kontrollvariable aufgrund des *„beauty is good" stereotype*), die die Studierenden den Dozierenden zuschrieben, auf das Erleben der Dozierenden wurde mit Regressionsanalysen mit Bootstrap überprüft. Die Ergebnisse sind in Tabelle 16.10 dargestellt. Insgesamt konnte regressionsanalytisch nur wenig Varianz aufgeklärt werden ($-.063 \leq$ korrigiertes $R^2 \leq .122$). Der mittlere Kompetenzeindruck zeigte sich als signifikanter Regressor bei den Dozierenden für das Erleben positiver Emotionen ($\beta = .366$, $p_{bt} = .029$), der eigenen Kompetenz in der Lehrveranstaltungssitzung ($\beta = .375$, $p_{bt} = .023$), von Flow ($\beta = .401$, $p_{bt} = .033$) sowie von Störungsmissmanagement ($\beta = -.399$, $p_{bt} = .028$). Auch wenn die Effektstärke für den Zusammenhang des Erlebens von Störungen in der Lehrveranstaltungssitzung mit dem mittleren, ersten Kompetenzeindruck der Studierenden verhältnismäßig hoch ausfiel, verfehlte es in der Analyse mit Bootstrap das Signifikanzniveau ($\beta = -.487$, $p_{bt} = .112$). Folglich können die Hypothesen 1, 3, 4 und 7 des zweiten Hypothesenkomplexes auf Basis der erzielten Ergebnisse angenommen werden.

16.4 Zusammenfassung, Interpretationen und Limitationen

Die vierte Studie ging der Frage nach, welche Auswirkungen Lernendenkompetenzerwartungen an Lehrpersonen haben. Dafür wurde eine Feldstudie im Hochschulkontext (Technische Universität Braunschweig) durchgeführt, die sowohl die Effekte auf die Lernenden selbst als auch auf die Lehrpersonen betrachtete.

16.4.1 Effekte auf die Lernenden

Die Intraklassenkorrelationen weisen darauf hin, dass die Wahrnehmung der Studierenden – sowohl der Lehrperson als auch des Unterrichtsgeschehens – weitestgehend geteilt wird (vgl. Geiser, 2010). Somit scheinen Studierende innerhalb eines Kurses ähnliche Standards der Bewertung heranzuziehen. Weitergehend weist die geteilte Wahrnehmung darauf hin, dass es ein Auftreten der Lehrpersonen gibt, das allgemein als kompetent beurteilt wird. Eine genauere Analyse,

Tabelle 16.10 Ergebnisse der linearen Regressionsanalysen mit Kontrollvariablen für das Dozierenden-Erleben (Studie 4)

Variable	Positive Emotionen					Negative Emotionen				
			Bootstrap[a]					Bootstrap[a]		
			95 %–CI					95 %–CI		
	β/*B*[b]	*p*	*LL*	*UL*	p_{bt}	β/*B*[b]	*p*	*LL*	*UL*	p_{bt}
Konstante	−0.000	1.000	−0.286	0.287	1.000	0.000	1.000	−0.309	0.295	1.000
Kompetenz[c]	.366	.033	0.045	0.684	.029	−.279	.122	−0.594	0.061	.108
Wärme[c]	−.215	.215	−0.575	0.109	.220	.445	.021	0.095	0.821	.025
Attraktivität[c]	.275	.088	−0.094	0.614	.139	−.092	.587	−0.460	0.301	.651
Modell	$R^2 = .171$, korr. $R^2 = .102$; $F(3, 36) = 2.470$, $p = .078$					$R^2 = .161$, korr. $R^2 = .084$; $F(3, 33) = 2.105$, $p = .118$				
Variable	Kompetenz DLV[d]					Wärme DLV[d]				
			Bootstrap[a]					Bootstrap[a]		
			95 %–CI					95 %–CI		
	β/*B*[b]	*p*	*LL*	*UL*	p_{bt}	β/*B*[b]	*p*	*LL*	*UL*	p_{bt}
Konstante	−0.000	1.000	−0.305	0.293	1.000	−0.000	1.000	−0.307	0.285	1.000
Kompetenz[c]	.375	.037	0.090	0.729	.023	.053	.753	−0.275	0.438	.740
Wärme[c]	−.185	.307	−0.593	0.158	.339	.381	.033	0.055	0.718	.032
Attraktivität[c]	.107	.509	−0.182	0.372	.441	.078	.619	−0.291	0.365	.631
Modell	$R^2 = .125$, korr. $R^2 = .054$; $F(3, 37) = 1.761$, $p = .171$					$R^2 = .189$, korr. $R^2 = .122$; $F(3, 36) = 2.799$, $p = .054$				

(Fortsetzung)

Tabelle 16.10 (Fortsetzung)

Variable	Flowerleben					Störungen				
			Bootstrap[a]					Bootstrap[a]		
			95 %–CI					95 %–CI		
	β/B^b	p	LL	UL	p_{bt}	β/B^b	p	LL	UL	p_{bt}
Konstante	−0.000	1.000	−0.281	0.304	1.000	0.000	1.000	−0.379	0.330	1.000
Kompetenz[c]	.401	.024	−0.057	0.700	.033	−.487	.027	−0.885	0.202	.112
Wärme[c]	−.124	.484	−0.459	0.207	.457	.225	.291	−0.217	0.507	.194
Attraktivität[c]	.203	.207	−0.108	0.478	.173	.023	.901	−0.340	0.452	.907
Modell	$R^2 = .184$, korr. $R^2 = .116$; $F(3, 36) = 2.700$, $p = .060$					$R^2 = .184$, korr. $R^2 = .086$; $F(3, 25) = 1.883$, $p = .158$				
Variable	Teilabsentismus					Störungsmissmanagement				
			Bootstrap[a]					Bootstrap[a]		
			95 %–CI					95 %–CI		
	β/B^b	p	LL	UL	p_{bt}	β/B^b	p	LL	UL	p_{bt}
Konstante	0.000	1.000	−0.302	0.354	1.000	−0.000	1.000	−0.344	0.331	1.000
Kompetenz[c]	.075	.689	−0.371	0.293	.612	−.399	.040	−0.747	−0.042	.028
Wärme[c]	.045	.816	−0.438	0.383	.818	.112	.567	−0.373	0.585	.644
Attraktivität[c]	−.122	.485	−0.467	0.278	.487	.186	.311	−0.133	0.509	.228
Modell	$R^2 = .019$, korr. $R^2 = -.063$; $F(3, 36) = 0.235$, $p = .871$					$R^2 = .180$, korr. $R^2 = .092$; $F(3, 28) = 2.051$, $p = .129$				

(Fortsetzung)

Tabelle 16.10 (Fortsetzung)

Variable	Beziehung				
			Bootstrap[a]		
			95 %–CI		
	β/B[b]	p	LL	UL	p_{bt}
Konstante	−0.000	1.000	−0.373	0.333	1.000
Kompetenz[c]	.024	.908	−0.367	0.570	.925
Wärme[c]	.149	.486	−0.317	0.679	.567
Attraktivität[c]	.088	.648	−0.323	0.518	.677
Modell	$R^2 = .042$, korr. $R^2 = -.057$; $F(3, 29) = 0.427$, $p = .735$				

Anmerkungen. Toleranz (Minimum) ≥ .736. Gesamtstichprobe: $N = 41$. Korr. R^2 = korrigiertes R^2. [a] Bootstrap: Perzentile; $N = 10000$. [b] Der standardisierte Regressionskoeffizient β entspricht aufgrund der vorherigen z-Transformierung aller Variablen dem unstandardisierten Koeffizienten B; für die Konstanten und die Konfidenzintervalle ist B angegeben. [c] Subjektive Beurteilung der Studierenden; Kompetenz- und Wärmebeurteilung beim ersten Eindruck (aggregierte Gruppenwerte). [d] Subjektive Beurteilung der Dozierenden von sich selbst während der Lehrveranstaltungssitzung.

welche Eigenschaften bzw. Eigenschaftskombinationen dies sind, müsste künftig untersucht werden, um handlungspraktische Ableitungen zu treffen. So kann eine zugewandte, offene Körpersprache von Lehrpersonen mehr Kompetenz ausstrahlen (Barmaki, 2014; Feldman & Prohaska, 1979; Gorham et al., 1997, 1999). Auch könnte die Stimmhöhe einen Einfluss haben, die Studien hierzu sind jedoch uneindeutig (Ko et al., 2009; Krahé & Papakonstantinou, 2020; Oleszikiewicz et al., 2017). Nichtsdestotrotz scheinen die Befunde darauf hinzudeuten, dass Kompetenz durch eine tiefe Stimme vermittelt wird sowie Wärme/Sympathie durch eine hohe Stimme. Ein Training zur gezielten Stimmmodulation kann deswegen in Lehrberufen sinnvoll sein. Weitergehend können auch die Sprechrate, die Nutzung von Füllwörtern und -lauten sowie die Betonung bedeutsam sein, um Kompetenz und Wärme zu vermitteln (vgl. Kiese-Himmel, 2016).

Die Zusammenhänge zwischen dem ersten Eindruck und der Wahrnehmung der Lehrveranstaltungssitzung folgten – unter Berücksichtigung von Kontrollvariablen – einem bestimmten Muster: Die motivationalen Variablen (Interesse, Langeweile) sowie die Wahrnehmung von Störungen und Störungsmissmanagement standen im Zusammenhang mit dem ersten Kompetenzeindruck. Der erste Wärmeeindruck hingegen war ein bedeutsamer Regressand für das emotionale Erleben (Freude, Frustration) sowie für die Wahrnehmung der Beziehung. Die Zusammenhänge zeigten sich fast ausschließlich auf der Gruppenebene. Die Attraktivitätsbeurteilung spielte unter Beachtung des Kompetenz- und des Wärmeeindrucks keine bedeutsame Rolle. Der hohe Zusammenhang von Kompetenzbeurteilung beim ersten Eindruck sowie innerhalb der Lehrveranstaltungssitzung ($r_S = .759$, $p < .001$) kann zu großen Teilen daran liegen, dass die identische Adjektivliste genutzt wurde. Gleiches gilt für die Wärmebeurteilung ($r_S = .815$, $p < .001$). Außerdem ist der zeitliche Abstand zwischen erstem Eindruck und der betrachteten Lehrveranstaltung im Mittel gering (ca. ein halbes Jahr), sodass große Abweichungen überraschend gewesen wären.

Zusammenfassend zeigten sich Kompetenz und Wärme als distinkte, aber jeweils wichtige Dimensionen in der sozialen Wahrnehmung im Lehr-Lern-Kontext, wie es im Rahmen des *stereotype content model* (s. Kapitel 5) postuliert wird. Die Erwartungseffekte auf die erwartungshaltende Person müssen also je nach Art der Erwartung unterschiedlich betrachtet werden.

16.4.2 Effekte auf die Lehrpersonen

Der Einfluss des ersten Lernendenkompetenzeindrucks – als Maß für Lernendenkompetenzerwartungen – auf die Lehrpersonen war vergleichsweise geringer

als die Wirkung auf die Lernenden selbst. So hing der mittlere studentische Kompetenzeindruck in den multivariaten Regressionsanalysen mit kleinen bis mittelgroßen Effekten positiv mit dem Erleben der eigenen Kompetenz in der Lehrveranstaltung, dem Erleben positiver Emotionen und von Flow sowie negativ mit dem Erleben von eigenem Störungsmissmanagement zusammen.

Die wenigen und teilweise kleinen Effekte auf das Erleben der Dozierenden könnten darin begründet liegen, dass sich innerhalb der zeitlich geringen Interaktion zwischen Lehrpersonen und Lernenden keine systematischen, interaktionalen Muster ausgebildet haben (interaktionale Schemata; Thies, 2010) – konträr zu Jussim und Harber (2005) sowie Raudenbusch (1984). Es könnte sein, dass am Ende des jeweiligen Semesters die Effekte auf die Lehrpersonen größer ausgefallen wären. So variiert die Effektstärke auch bei Rosenthal und Jacobson (1971), mit einem Höhepunkt ein Jahr nach der Kompetenzmanipulation.

Weitergehend kann vermutet werden, dass Lehrpersonen im Verlauf der Lehrpraxis Coping-Strategien aufbauen und sich somit weniger bzw. weniger schnell durch negatives Verhalten der Lernenden beeinflussen lassen. Ob Coping-Strategien bedeutsam sind, müsste in weitergehenden Untersuchungen betrachtet werden. Andererseits kann es auch sein, dass Lehrpersonen an einer Hochschule mit geringem Ausmaß an Lehrverpflichtungen (und weiteren Aufgaben z. B. in der Forschung) als Lehrkräfte an allgemeinbildenden Schulen und dem damit einhergehenden Selbstkonzept (geringere Identifikation als Lehrende) weniger durch problematische Lehr-Lern-Situationen beeinflusst werden und/oder weniger sensibel dafür sind.

Die erzielten Effektstärken liegen unterhalb derer, die Feldman und Prohaska (1979) berichteten ($|d|$ bis zu 1.93). Die kleinen Effektstärken des mittleren ersten Lernendenkompetenzeindrucks von den Lehrpersonen sind hier aber vor dem Hintergrund der Feldstudie zu interpretieren. Das eigene Erleben und das der Lehrveranstaltungssitzung können von vielen weiteren Einflüssen beeinträchtigt sein, z. B. vom Wetter, von der Tageszeit, vom erlebten Stress außerhalb der Lehrsituation, die innerhalb einer spezifischen Lehrveranstaltungssitzung stärker wirksam sein könnten. Deswegen braucht es größere Stichproben (s. u.) bzw. auch längsschnittliche Untersuchungen. Weitergehend kann die Höhe der Effektstärke auch davon abhängig sein, ob der Kompetenzeindruck bzw. die -erwartungen von den Lernenden in größerem oder kleinerem Ausmaß geteilt wird (vgl. Jamieson et al., 1987). Eine diesbezügliche statistische Prüfung konnte hier aufgrund der kleinen Stichprobe ($N = 41$) nicht sinnvoll durchgeführt werden. Zukünftige Studien im Feld zu Lernendenkompetenzerwartungen sollten dies in der Untersuchungsplanung berücksichtigen und von kleinen Effektstärken ausgehen.

16.4.3 Limitationen

Die bedeutsamste Limitation innerhalb der Untersuchung wurde bereits eingangs bei der Beschreibung des Vorgehens beschrieben. Sie besteht darin, dass der erste Eindruck der Studierenden von der dozierenden Person retrospektiv erfasst wurde. Durch diese Methode kann es sein, dass nicht der tatsächliche erste Eindruck erfasst wurde. Auch wenn die Methode Schwächen aufweist, kann sie unter der Annahme, dass der erste Eindruck stabil repräsentiert ist (vgl. Schweer, 2019), als angemessen bewertet werden, insbesondere im Kontext einer Feldstudie. Weitergehend zeigten die Korrelationen der Skalen Kompetenz und Wärme beim ersten Eindruck mit der jeweiligen Beurteilung innerhalb der Lehrveranstaltung keine perfekten Zusammenhänge. Dies kann als Indiz dafür gedeutet werden, dass die teilnehmenden Studierenden ihren tatsächlichen ersten Eindruck angaben. Weitergehend lag die erste Begegnung mit der dozierenden Person, für die der erste Eindruck erinnert werden musste, im Mittel nur etwa ein halbes Jahr zurück, was als überschaubarer Zeitraum gewertet werden kann. Aufgrund dieses „pseudo-längsschnittlichen" Ansatzes sowie der ausschließlichen Betrachtung von Zusammenhängen lassen sich aber keine Kausalitäten ableiten (vgl. Döring & Bortz, 2016). Perspektivisch könnten Untersuchungen in Blockseminaren, bei externen Lehrbeauftragten sowie in der Erwachsenenbildung durchgeführt werden, weil dort kontrollierter der erste Eindruck erfragt werden kann. Im schulischen Kontext ist dies beispielsweise beim Übergang in weiterführende Schulen möglich.

Wie in den vorherigen Studien wurde auch hier auf eine Gelegenheitsstichprobe zurückgegriffen. Weitergehend gab es keine repräsentative Auswahl an Lehrveranstaltungen, sodass Veranstaltungen aus den Geistes- und Erziehungswissenschaften stärker repräsentiert waren. Dies und die freiwillige Teilnahme der Dozierenden können eine Erklärung für die Deckeneffekte und die damit besonders positive Beurteilung sein. So könnten Dozierende insbesondere in den Erziehungswissenschaften besonders sensibel für die Gestaltung von Lehre sein und ein entsprechendes Auftreten entwickelt haben (im Sinne von positiv zu bewertendem Klassenführungsverhalten). Weitergehend kann angenommen werden, dass insbesondere Dozierende, die besonders gute Lehrveranstaltungen anbieten (z. B. rückgemeldet über regelmäßige Evaluationen), an der Untersuchung teilnahmen. Dadurch kann die Motivation an der Teilnahme an einer Studie zur Klassenführung (Cover-Story) erhöht sein. Dies schränkt allerdings die Generalisierbarkeit der Ergebnisse deutlich ein, auch weil davon ausgegangen werden muss, dass die tatsächliche Varianz der Variablen in der Untersuchung nicht

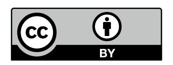
korrekt repräsentiert ist – was auch zu den kleinen Effekten auf die Lehrpersonen führen könnte. Nichtsdestotrotz kann davon ausgegangen werden, dass die Richtung der Zusammenhänge (positiv oder negativ) der Variablen angemessen beschrieben wird.

Eine weitere Problematik besteht in Bezug auf die kleine Zahl der teilnehmenden Dozierenden ($N = 41$) resp. der untersuchten Lehrveranstaltungen ($N = 42$). Dadurch ergab sich eine reduzierte statistische Power (vgl. Hager, 2004), insbesondere bei der Betrachtung auf Gruppenebene bzw. der Lehrpersonen. Dies ist forschungsökonomisch begründet, da die Erhebungen sehr aufwendig waren (zwei Messzeitpunkte innerhalb einer Lehrveranstaltungssitzung), und könnte künftig um weitere Veranstaltungen ergänzt werden.

Eine zusätzliche Limitation ergibt sich durch die verwendeten Fragebögen. Viele der verwendeten Items wurden ursprünglich für den Schulkontext erstellt und für die Hochschullehre sprachlich angepasst. Auch wenn eine intensive Itemanalyse durchgeführt sowie die faktorielle Validität überprüft wurde, hätten andere Fragebögen, die spezifisch die Hochschullehre fokussieren bzw. Erwachsene adressieren, Anwendung finden können: z. B. Münsteraner Fragebogen zur Evaluation von Seminaren (MFE-Sr) bzw. von Vorlesungen (MFE-Vr) von Thielsch & Hirschfeld (2014a, 2014b) oder eine lehrveranstaltungsspezifische Version des Fragebogens zur Lern- und Leistungsmotivation im Studium (LLMS) von Sudman et al. (2014).

Demgegenüber wurden unterschiedliche Formate von Lehrveranstaltungen betrachtet, von kleinen Seminaren (mit $n = 7$ Teilnehmenden) bis zu größeren Vorlesungen (mit $n = 78$ Teilnehmenden). Außerdem wurden unterschiedliche Fachdisziplinen betrachtet. Auch wenn Lehrveranstaltungen der Geistes- und Erziehungswissenschaften am häufigsten vertreten waren, sind sie dennoch sehr heterogen. Nichtsdestotrotz sollten in weiteren Untersuchungen auch stärker andere Fachdisziplinen, z. B. die Natur- und Ingenieurswissenschaften, fokussiert werden.

Positiv an der Untersuchung hervorzuheben ist die Anzahl der teilnehmenden Studierenden ($N = 663$ Ersteilnahmen). Dadurch ergab sich für die Berechnungen der Zusammenhänge auf der Lernendenseite eine hohe statistische Power. Insgesamt kann diese Feldstudie weitere Erkenntnisse zur Wirkung von Lernendenerwartungen auf Lernende liefern sowie die rare Befundlage zur Wirkung derselben auf Lehrpersonen erweitern.


BY

Teil VI

DISKUSSION

17 Zusammenfassende Diskussion der Ergebnisse

Ausgehend vom transaktionalen Modell der Lehr-Lern-Interaktion von Nickel (1985), wurde der in der (Pädagogischen) Psychologie weitestgehend vernachlässigten Frage nachgegangen, wie Lernendenkompetenzerwartungen an Lehrpersonen wirken. Lehrendenerwartungen zeigen sich als ein interaktionaler Baustein, der das Angebot der Lehrpersonen an Lernende beeinflusst (Angebots-Nutzungs-Modell; u. a. Helmke, 2017; Reusser & Pauli, 2010). Besonders wenig ist über die Wirkung der Lernendenkompetenzerwartungen, insbesondere auf Lehrpersonen, bekannt. In anderen Worten: Es gibt kaum empirische Evidenz, die auf selbsterfüllende bzw. -erhaltende Prophezeiungen von Lernendenerwartungen in der Lehr-Lern-Interaktion hinweist. In Anlehnung an das Modell von Nickel (1985) wurden deswegen folgende drei Fragen fokussiert:

1. Wie stark werden Lernendenkompetenzerwartungen an Lehrpersonen durch (Geschlechts-)Stereotype bestimmt?
2. Wie wird das nonverbale Verhalten von Lernenden in Abhängigkeit von den Lernendenkompetenzerwartungen an Lehrpersonen beeinflusst?
3. Wie wirken initiale Lernendenkompetenzerwartungen an Lehrpersonen auf die Wahrnehmung der Lehr-Lern-Situation – sowohl bei Lernenden als auch bei Lehrpersonen?

Die erzielten Befunde der vier durchgeführten Studien sind in Abbildung 17.1 kurz und überblicksartig festgehalten.

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_17

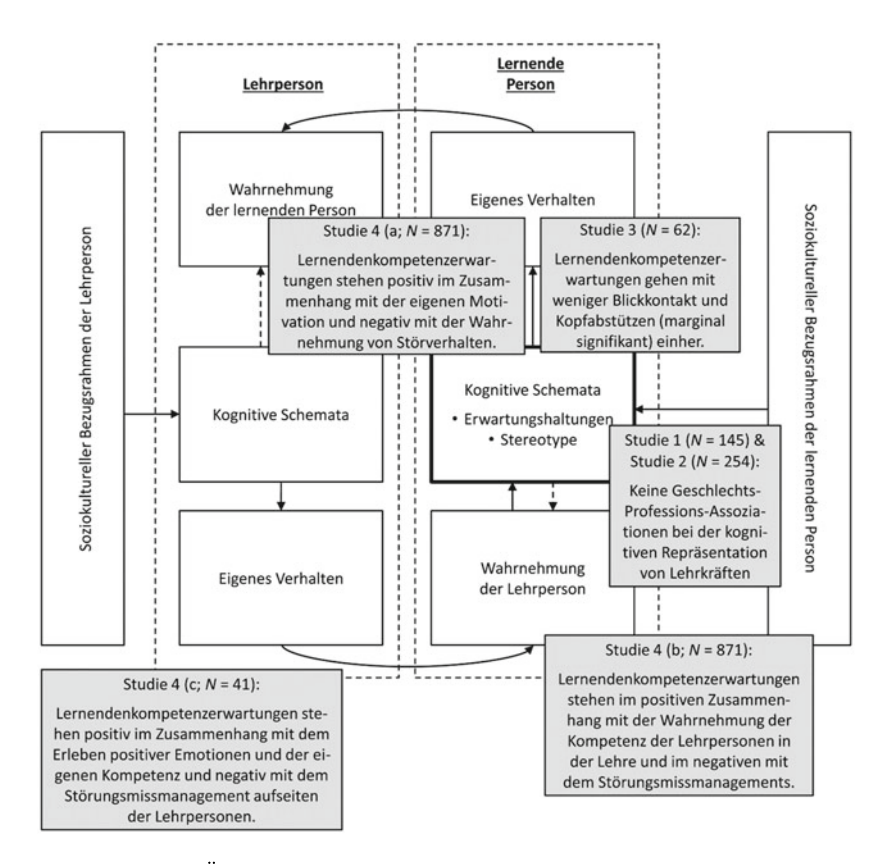


Abbildung 17.1 Überblick über die erzielten Ergebnisse mit Bezug auf die Lernendenkompetenzerwartungen

Hinweise auf Geschlechts-Professions-Assoziationen (Guimond & Roussel, 2001; Miller et al., 2018; Nosek et al. 2009; Plante et al., 2009; Smyth & Nosek, 2015; Su et al., 2009) zeigten sich bisher für Lehrkräfte im Unterrichtsfachvergleich Deutsch und Physik (Degner et al., 2019; Zander et al., 2015). Die Studien 1 und 2 fokussierten solche Geschlechts-Professions-Assoziationen, abweichend zu Degner et al. (2019) und Zander et al. (2015) wurden die Unterrichtsfächer Deutsch und Mathematik herangezogen. Beide Studien nutzten Fallvignetten, in denen eine Lehrperson beschrieben wurde. Die Hypothesen, dass weibliche Lehrkräfte im Deutschunterricht kompetenter und im Mathematikunterricht weniger

kompetent beurteilt werden als männliche, fanden auf Basis der erzielten Ergebnisse unter Berücksichtigung von Kontrollvariablen weder in Studie 1 noch in Studie 2 Unterstützung. Weitergehend konnte Studie 2 auch nicht aufzeigen, dass weibliche Lehrkräfte in der Grundschule, in der Erziehung ein wesentlicher Aspekt der Tätigkeit ist (relativ zum Gymnasium), als kompetenter beurteilt werden als männliche Lehrkräfte in der Grundschule oder weibliche im Gymnasium. Weibliche Lehrkräfte wurden in Studie 1 – unter Kontrolle von Wärme- (*stereotype content model*; s. Kapitel 5) und Attraktivitätsbeurteilung (*„beauty is good" stereotype*; Dion et al., 1972; Eagly et al., 1991; Langlois et al., 2000) – unabhängig vom Fach kompetenter beurteilt. Der Effekt war jedoch klein (Cohen, 1988). Auch in Studie 2 konnte ein kleiner Effekt dafür ausgemacht werden, dass weibliche Lehrkräfte kompetenter beurteilt werden als männliche. Dieser Effekt verschwindet jedoch unter Berücksichtigung der Wärmebeurteilung. Somit müssen die Hypothesen zu den Geschlechts-Professions-Assoziationen verworfen werden. Dies kann an den ausgewählten Unterrichtsfächern liegen: Es kann vermutet werden, dass Physik (im Vergleich zu Mathematik) als besonders männliche Domäne wahrgenommen wird, Deutsch jedoch nicht als besonders weibliche. Es ist außerdem denkbar, dass männliche Lehrkräfte als Substereotyp (vgl. Machunsky, 2020) repräsentiert sind, also nicht dem Stereotyp eines („typischen" bzw. „durchschnittlichen") Mannes entsprechen, weil sie einen weiblich assoziierten Beruf ergriffen haben und somit allgemein wärmer und weniger kompetent beschrieben werden. Welche (Unterrichts-)Fächer mit welchem Geschlecht assoziiert sind, sollte in Folgestudien spezifisch für den schulischen bzw. den jeweiligen Kontext untersucht werden, sowohl mit expliziten – wie in Studie 1 und 2 sowie bei Zander et al. (2015) – als auch mit (eher) impliziten Maßen – wie der *reverse correlation classification task* bei Degner et al. (2019) oder mit Impliziten Assoziationstests (vgl. Eckes, 2020). Insgesamt spiegeln die Ergebnisse der Studien 1 und 2 die bisherigen unsystematischen Befunde wider (s. Kapitel 11).

Die zweite Fragestellung zielte auf das nonverbale Verhalten. Es wurde davon ausgegangen, dass sich die Kompetenzerwartungen im nonverbalen Verhalten widerspiegeln. Deswegen sollten sich Lernende mit unterschiedlichen Kompetenzerwartungen an Lehrpersonen auch unterschiedlich verhalten. Einige bisherige Studien zeigten, dass Lernende, die eine kompetente Lehrkraft erwarteten, sich zugewandter und weniger störend verhielten (Feldman & Prohaska, 1979; Jamieson et al., 1987) sowie höhere Leistungen erzielten (Adediwura & Tayo, 2007; Edwards et al., 2009; Feldman & Prohaska, 1979; Jamieson et al., 1987). Diese Annahme kann auf Basis der Ergebnisse der Studie 3 verworfen werden. Nichtsdestotrotz zeigten sich kleine bis mittelgroße Effekte für den Blickkontakt sowie für das Abstützen des Kopfes: Proband*innen in der Galatea-Bedingung, also

bei der Erwartung einer kompetenten Lehrperson, hielten weniger Blickkontakt und sahen häufiger weg als Proband*innen in der Golem-Bedingung. Außerdem zeigte sich ein statistischer Trend dafür, dass Proband*innen in der Galatea-Bedingung weniger lang ihren Kopf abstützten. Somit verhielten sie sich weniger zugewandt, was im Widerspruch zu bisherigen Studienbefunden steht. Der Stimulus, der in der dritten Studie Anwendung fand, war ein relativ monotoner Vortrag. Bei Feldman und Prohaska (1979) gab es dyadische Interaktionen, bei Jamieson et al. (1987) ein „natürliches" Lehr-Lern-Setting im Feld. Somit kann argumentiert werden, dass Lernende sich nicht per se zugewandter verhalten, wenn sie eine kompetente Lehrperson erwarten, sondern mehr „Haltung bewahren" (im Sinne von mehr Körperspannung). Bei wenig motivierenden und stimulierenden Reizen vonseiten der Lehrperson versuchen Lernende möglicherweise, sich Stimulation außerhalb des Lehr-Lern-Settings zu holen, sodass sie häufiger weggucken. Demgegenüber verlieren Lernende, die eine weniger kompetente Lehrperson erwarten, die Körperspannung: Sie stützen sich ab, z. B. ihren Kopf auf den Armen, was ihnen einen längeren Blickkontakt ermöglicht. Dies ist ein Verhalten, das bei Mehrabian (1969) als Entspanntheit (*relaxation*) beschrieben wird und mit Status und Macht assoziiert ist. Es kann also vermutet werden, dass geringe Kompetenzerwartungen über die Entspanntheit weniger Respekt vor der Lehrperson vermitteln. Somit müssen sowohl die Zugewandtheit (*immediacy*) und die Entspanntheit (*relaxation*) gemeinsam berücksichtigt werden und sind je unterschiedlich für verschiedene Lehr-Lern-Settings bedeutsam. Die erzielten Ergebnisse sprechen für zwei „Langeweileprofile", die mit Entspanntheit in Beziehung stehen. Diese Interpretation bedarf jedoch künftiger Forschung.

Die finale dritte Fragestellung, wie initiale Lernendenkompetenzerwartungen an Lehrpersonen auf die Wahrnehmung von Lehre wirken, wurde in der Studie 4 in einer Feldstudie untersucht. Dabei ist die Wirkung auf die Lernenden selbst von der Wirkung auf die Lehrperson zu unterscheiden. Die (retrospektiv erfassten) Lernendenkompetenzerwartungen (erster Eindruck) gingen auf Lernendenseite mit Motivation sowie mit der Störungswahrnehmung und der Wahrnehmung von Störungsmissmanagement einher, wenn für die Attraktivitätsbeurteilung und die (retrospektiv erfassten) Wärmeerwartungen (erster Eindruck) kontrolliert wird: Je höher die Lernendenkompetenzerwartungen waren, desto mehr Interesse und weniger Langeweile wurden berichtet, auch wurden weniger Störungen sowie ein besseres Störungsmanagement wahrgenommen. Die Wärmeerwartung war demgegenüber insbesondere im Zusammenhang mit den sozioemotionalen Variablen (Freude, Frustration und Beziehungswahrnehmung) innerhalb einer Lehrveranstaltung. Dieses Muster zeigt die Bedeutsamkeit von

Erwartungen an unterschiedliche Bereiche resp. psychologische Wahrnehmungsdimensionen. Die Effekte von Kompetenzerwartungen sind nicht mit denen von Wärmeerwartungen vergleichbar. Die hier durchgeführte Differenzierung basiert auf dem *stereotype content model* (u. a. Fiske et al., 1999; s. Kapitel 5), alternative Unterscheidungen sind jedoch möglich (z. B. Koch et al., 2016). Nichtsdestotrotz ergänzen diese Ergebnisse bisherige Befunde, die ausschließlich die Lernendenkompetenzerwartungen berücksichtigen (Feldman & Prohaska, 1979; Feldman & Theiss, 1982; bei Jamieson et al., 1987, wird auch die Motivierungsfähigkeit unter Kompetenz subsumiert). Zu beachten ist, dass sich die Zusammenhänge der vierten Studie in den Mehrebenenanalysen fast ausschließlich auf der Gruppenebene (*between*) zeigten, die Individualebene (*within*) kann jedoch nur wenig Varianz aufklären. In der Forschung und auch in der Theorie zu (Kompetenz-) Erwartungen werden gruppendynamische Prozesse bei selbsterfüllenden Prophezeiungen nicht berücksichtigt. Die Befunde aus Studie 4 zeigen den Bedarf, solche Prozesse systematisch zu betrachten.

Die Wirkung der Lernendenkompetenzerwartungen auf die Lehrpersonen war im Vergleich zur Wirkung auf die Lernenden gering und zeigte keine eindeutige Systematik. Der Befund von Feldman und Prohaska (1979), dass sich vermeintlich kompetente Lehrpersonen aus Studierendensicht auch als kompetenter beschreiben, konnte repliziert werden. Unter Berücksichtigung der retrospektiv erfassten, initialen Wärmebeurteilung (erster Eindruck) und der Attraktivitätsbeurteilung zeigte sich der mittlere Kompetenzeindruck im positiven Zusammenhang mit dem Erleben positiver Emotionen, von Flow und mit der eigenen Kompetenz in einer Lehrveranstaltungssitzung sowie im negativen Zusammenhang mit dem Erleben von Störungsmissmanagement. Somit lässt sich aus den erzielten Ergebnissen schließen, dass auch Lehrpersonen von hohen Lernendenkompetenzerwartungen an sie profitieren. Die eher kleinen bis mittelgroßen Zusammenhänge ($.366 \leq |\beta| \leq .401$) sind in Anbetracht der Feldstudie und damit aufgrund vieler unterschiedlicher Beeinflussungen auf die Einschätzung einer spezifischen Lehrveranstaltungssitzung als bedeutsam einzuordnen. Außerdem ist es möglich, dass Lehrpersonen durch höhere Machtmittel in der asymmetrischen Lehr-Lern-Interaktion (u. a. Misamer, 2019; Steins, 2013; Thies, 2010) weniger in ihrem Erleben und Handeln durch die Lernenden beeinflusst werden als Lernende durch Lehrpersonen. Somit können also keine großen Effekte erwartet werden. Insgesamt kann resümiert werden, dass Lernendenkompetenzerwartungen die Lehr-Lern-Interaktion beeinflussen, und es braucht weitere Befunde, um die konkreten Auswirkungen auszumachen.


BY



18 Limitationen und Stärken

Studienspezifische Limitationen sind bereits in den jeweiligen Kapiteln diskutiert worden. Im Folgenden sollen übergreifende Limitationen aufgezeigt werden, die die Interpretation der Ergebnisse beeinflussen. Hier sind insbesondere sechs Aspekte zu berücksichtigen: (1.) Die ersten drei Studien fokussieren Lehrkräfte in allgemeinbildenden Schulen und die vierte Studie wurde im Hochschulkontext durchgeführt. Auch wenn davon ausgegangen werden kann, dass die allgemeinen Prozesse des Lehrens und Lernens, inklusive der Lehr-Lern-Interaktion, ähnlich verlaufen (vgl. Klein, 2011), sind die erzielten Ergebnisse nicht ohne weitere Überprüfung auf den jeweils anderen Kontext (oder weitere wie der Erwachsenenbildung z. B. in Volkshochschulen) übertragbar. (2.) Ein zweiter Aspekt, der die Interpretation der Ergebnisse limitiert, liegt in der Stichprobenzusammensetzung. Die Teilnehmenden aller Studien waren Erwachsene. Es ist zwar davon auszugehen, dass die kognitiven Repräsentationen und (Geschlechts-)Stereotype innerhalb eines kulturellen Kontexts zwischen Erwachsenen, Kindern (ab 6 Jahren; insbesondere ab 10 Jahren) und Jugendlichen nicht wesentlich verschieden sind (vgl. Beelmann & Neudecker, 2020; Raabe & Beelmann, 2009), eine Untersuchung der Fragestellungen mit Kindern und Jugendlichen ist aber dennoch wünschenswert, um die Befunde zu unterstützen oder zu widerlegen und um ein differenziertes Bild zu erhalten (z. B. variierende Effektstärken). (3.) Ein weiterer, wesentlicher Kritikpunkt an der Methodik bezieht sich auf die Stichprobenziehung. Alle Studien wurden mit Gelegenheitsstichproben (bzw. Ad-hoc-Stichproben; Döring & Bortz, 2016) mit einem allgemeinen Teilnahmeaufruf durchgeführt. Diese Form der Stichprobenziehung beinhaltet die Gefahr von Selbstselektions-Effekten und ist selten repräsentativ für die Gesamtbevölkerung resp. die Zielpopulation. Dies sollte für die Studien 1 und 2 weniger bedeutsam sein, da Stereotype sozial geteilte Wissensstrukturen sind (vgl. Cuddy et al., 2008; Stangor, 2009). Unter Berücksichtigung der Teilnehmendenstruktur kann davon ausgegangen werden,

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_18

dass die Ergebnisse zu den Stereotypen eher konservativer (also mit kleineren Effektenstärken) ausfielen, da hauptsächlich Proband*innen aus psycho-sozialen Berufen oder Studienfächern teilnahmen. Diese Teilnehmenden sollten aufgrund ihrer Ausbildung sensibler gegenüber Stereotypen sein (so geht etwa der fundamentale Attributionsfehler mit weniger sozialpsychologischer Kompetenz einher; Gollwitzer & Bargh, 2018). Analoge Überlegungen lassen sich für Studie 3 aufstellen. Besonders problematisch zeigte sich die Gelegenheitsstichprobe in Studie 4 bei den Dozierenden. In Anbetracht der besonders positiven Bewertungen der Lehrpersonen muss von einem Selektionseffekt ausgegangen werden, der die beobachtete Varianz einschränkt, jedoch die Richtung der aufgezeigten Zusammenhänge nicht beeinflussen sollte. (4.) Eine weitere Limitation betrifft die kleinen Stichprobengrößen, insbesondere in Studie 3 und die Lehrpersonen-Stichprobe in Studie 4 mit jeweils unter 70 Proband*innen. Trotz der Ziehung von Gelegenheitsstichproben war es schwer möglich, Personen zur Teilnahme an aufwendigeren Untersuchungen zu gewinnen. Dadurch reduziert sich die statistische Power deutlich, weswegen auch keine Anpassung des Signifikanzniveaus durchgeführt wurde – was die Power erneut reduziert und die Wahrscheinlichkeit für Beta-Fehler (Fehler 2. Art) erhöht hätte (vgl. Bortz & Schuster, 2010; Hager, 2004). Die Diskussion der Ergebnisse basierte deswegen auch stärker auf den erzielten Effektstärken. Um Verletzungen von Testvoraussetzungen zu begegnen und ausreißende Werte angemessen zu berücksichtigen, wurden die Ergebnisse zumeist mit Rangkorrelationen (nach Spearman) und Verfahren mit Bootstrap berechnet. (5.) Außerdem muss die Auswahl der Items beachtet werden. Die Items zur Erfassung der Kompetenz und Wärme sowie der Attraktivität wurden eigens für die Studien ausgewählt, ohne jedoch eine inhaltliche Validierung durchzuführen. Die durchgeführten Faktoren- und Reliabilitätsanalysen unterstützen jedoch die theoretischen Überlegungen, die den Itemkonstruktionen zugrunde lagen. Allerdings ist davon auszugehen, dass die Kompetenz- und Wärmebeurteilungen mit der verwendeten Likert-Skala subjektive Maße sind. Dies ist insbesondere für die Studien 1 und 2 problematisch, da dies zu einem *shifting standard* (Biernat et al., 1991; Foschi, 2000) geführt haben könnte und sich somit die Geschlechts-Professions-Assoziationen nicht einstellten, da die beschriebenen Personen in den Fallvignetten mit der jeweiligen Gruppe verglichen wurden (z. B. „für eine männliche Lehrkraft wirkt die Person ziemlich kompetent“). Es gibt also einen Bedarf der Replikation mit objektiven Maßen (z. B. Bewertung mit Punkten oder Einschätzung einer Erfolgswahrscheinlichkeit). Darüber hinaus wurden einige der verwendeten, bereits validierten Skalen umformuliert bzw. Erhebungsinstrumente angepasst, um sie an die Gegebenheiten der jeweiligen

Studie anzupassen. (6.) Das Ziel, die Wirkung von Lernendenkompetenzerwartungen auf Lehrpersonen zu untersuchen, wurde mit Zusammenhangsanalysen in der vierten Studie (Feldstudie) umgesetzt. Dadurch können keine Aussagen zu Kausalitäten getroffen werden. Eine Untersuchung, die längsschnittlich (im Hochschulkontext z. B. über ein Semester) mit mehreren Messzeitpunkten verläuft und mit dem Erfassen der ersten Kompetenzerwartung bzw. -beurteilung beginnt, wäre ein angemessenes Verfahren, um sowohl Kausalaussagen zu ermöglichen als auch kurz- und langfristige Effekte von Lernendenkompetenzerwartungen zu betrachten.

Neben diesen Limitationen sollen auch die wesentlichen Stärken der Studien erwähnt werden. (1.) Die Studien 1, 2 und 3 haben ein experimentelles Design und die Studien 1 und 2 mit randomisierter Zuordnung, sodass der Einfluss der Manipulationen kausal interpretiert werden kann. (2.) In Studie 3 standen keine individuellen subjektiven Angaben der Proband*innen im Fokus der Auswertung, sondern das nonverbale Verhalten der Proband*innen, was einerseits besonders aufwendig zu erheben ist und andererseits selten in der psychologischen Forschung berücksichtigt wird (Döring & Bortz, 2016; Weber, 2012). (3.) Einige Stichprobengrößen in den Studien sind relativ klein, allerdings ist die Stichprobengröße der Lernenden in Studie 4 mit 663 Erstteilnahmen (eindeutig identifizierbare Erstteilnahmen an der Studie, da Mehrfachteilnahmen möglich waren) und insgesamt 871 Angaben (inkl. Mehrfachteilnahmen) groß. (4.) Neben Feldman und Prohaska (1979) sowie Feldman und Theiss (1982) ist die Studie 4 die einzige (deutsch- und/oder englischsprachige) Untersuchung, die die Zusammenhänge von Lernendenkompetenzerwartungen an Lehrpersonen mit der Wahrnehmung der Lehrpersonen von sich und der eigenen Lehre fokussiert. Im Vergleich zu Feldman und Prohaska (1979) sowie Feldman und Theiss (1982) wurde zudem eine Feldstudie durchgeführt, was die externe Validität der Befunde erhöht (Döring und Bortz, 2016). (5.) Die vierte Studie berücksichtigte sowohl die Gruppen- als auch die Individualebene als Analyseeinheiten, um die Zusammenhänge der Lernendenkompetenzerwartungen auf die Lernenden selbst zu betrachten. Zumeist wird auf die unterschiedlichen Ebenen zu selbsterfüllenden Prophezeiungen im Lehr-Lern-Kontext in Forschung und Theorie kaum eingegangen. Die erzielten Ergebnisse weisen auf die Bedeutsamkeit der Unterscheidung hin, da die Gruppenebene auf Basis der erzielten Ergebnisse deutlich mehr Varianz aufklären kann als die Individualebene.


BY

19 Praktische und theoretische Ableitungen

Zusammengefasst weisen die Befunde der durchgeführten Studien auf einen Effekt der Lernendenkompetenzerwartungen an Lehrpersonen auf das Erleben und Verhalten von sowohl Lernenden als auch Lehrpersonen hin. Die Effektstärken und Zusammenhänge waren jedoch mehrheitlich klein, insbesondere in Bezug auf das nonverbale Verhalten der Lernenden und auch hinsichtlich des Erlebens der Lehrpersonen. Nichtsdestotrotz kann abgeleitet werden, dass hohe Lernendenkompetenzerwartungen im Zusammenhang mit positiven und erwünschten Lehr-Lern-Interaktionen stehen. Für Lehrpersonen ist es also förderlich, hohe Kompetenz auszustrahlen – sowohl für die Lernenden als auch für sie selbst. Im Folgenden werden zuerst Ableitungen beschrieben, wie Lehrpersonen kompetenter erscheinen und einen ersten Eindruck aktiv beeinflussen können, der selbsterfüllend wirken kann. Anschließend wird dies im Kontext der Lehr-Lern-Interaktion, insbesondere mit Fokus auf die Gestaltung der Klassenführung, betrachtet. Abschließend werden, ergänzend zum Modell von Nickel (1985), Erwartungseffekte in einem synergetisch-systemischen Ansatz der Lehr-Lern-Interaktion beschrieben.

19.1 Gestaltung des ersten Eindrucks zur Erhöhung der Lernendenkompetenzerwartungen an Lehrpersonen

Die erzielten Ergebnisse können insbesondere für die Gestaltung von kurzen Lehr-Lern-Interaktionen sowie bei beginnenden Interaktionen, die für Erwartungseffekte besonders vulnerabel sind (vgl. bei Lehrendenkompetenzerwartungen Jussim & Harber, 2005), praktisch genutzt werden. Insbesondere die Gestaltung des ersten Eindrucks, den die Lernenden von der Kompetenz der Lehrperson aufbauen, kann auf verschiedene Arten beeinflusst werden, um einen

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8_19

positiven Engelskreis in Gang zu setzen (vgl. Nickel et al., 1985) und um möglichen Geschlechts-Professions-Assoziationen zu begegnen (für Physik-Unterricht s. Degner et al., 2019; Zander et al., 2015; vgl. Studie 1 und Studie 2). Weitergehend kann die positive Anbahnung einer Kompetenzerwartung auch Eingang in die Konzipierung von Klassenführungstrainings (z. B. Hannemann, 2020; Kiel et al., 2013) finden, die theoretische und praktische Schnittstellen zu selbsterfüllenden Prophezeiungen bieten.

Kounin (2006, Original von 1976) konnte verschiedene Dimensionen erfolgreicher Klassenführung ausmachen, darunter Allgegenwärtigkeit und Überlappung. Dies beschreibt die Fähigkeit von Lehrpersonen, den Lernenden zu vermitteln, dass sie über die aktuelle Lehr-Lern-Situation informiert sind und eingreifen können. Dies wird nach Kiel et al. (2013) u. a. mit Präsenzmaßnahmen umgesetzt. Präsenzmaßnahmen sind Maßnahmen, die den Lernenden aufzeigen, dass die Lehrperson im ganzen Raum präsent ist, alles in der Lehr-Lern-Situation im Blick hat und bei Störungen eingreift (vgl. Haag, 2018; Kiel et al., 2013). Dazu gehören u. a. Körpersprache, darunter Blickkontakt, Bewegung im Raum und selbstsicheres Auftreten. Diese Maßnahmen können Kompetenz vermitteln – und interaktionale, kognitive Schemata (Thies, 2010) bei den Lernenden entstehen lassen – und weitergehend die Lehr-Lern-Interaktion nach dem Modell von Nickel (1985) zu einem positiven Engelskreislauf führen. Kiel et al. (2013) unterscheiden dabei nonverbale und verbale sowie intervenierende und präventive Maßnahmen. Der Lehrberuf ist einer der Berufe, „[…] bei denen die Mechanismen nonverbaler Kommunikation eine größere Rolle spielen“ (Kiel et al., 2013, S. 59), die auch Kompetenz ausstrahlen können.

Zum nonverbalen Verhalten beschreibt Mehrabian (1969), dass die Zugewandtheit (*immediacy*), die sich in Berührung, Distanz, Vorbeugen, Blickkontakt und Körperausrichtung manifestiert, Sympathie gegenüber der anderen Person resp. Personen der Kommunikation vermittelt, sodass mit diesen Verhaltensweisen der Wärmeeindruck vermittelt wird. Dafür ist jedoch eine entspannte Körperhaltung notwendig bzw. förderlich. In Ergänzung zur Zugewandtheit bildet die Entspanntheit der Körperhaltung (*relaxation*) einen Hinweis auf Macht- bzw. Statusunterschiede, die auch mit der Kompetenzwahrnehmung im Zusammenhang stehen können. Mehrabian (1969) beschreibt, dass eine mittlere Körperanspannung am wohlwollendsten von Kommunikationspartner*innen aufgenommen wird. Eine mittlere Körperspannung sollte von Lehrpersonen in Kombination mit einem zugewandten Verhalten gezeigt werden, um sowohl Wärme als auch Kompetenz auszustrahlen. Bisher zeigte sich zur Kompetenzwahrnehmung, dass eine zugewandte Körpersprache von Lehrpersonen mit der Wahrnehmung von höherer Kompetenz verknüpft ist (Feldman & Prohaska, 1979; Gorham et al., 1997,

1999). Eine offene, nicht geschlossene Haltung sowie Nähe hingen mit positiv bewerteter Lehrperformanz zusammen (Barmaki, 2014).

Zusätzlich zur Körperhaltung kann auch die Stimmhöhe einen Einfluss auf die Ausbildung von Kompetenzerwartungen haben. So beurteilten Proband*innen, die kein Wissen über die Fragestellung hatten, sowohl männliche als auch weibliche Stimmen, deren Ton technisch vertieft wurde, kompetenter und vertrauenswürdiger als Stimmen, deren Ton technisch erhöht wurde. Allerdings wurden technisch erhöhte Frauenstimmen wärmer bzw. sympathischer wahrgenommen als technisch vertiefte Frauenstimmen (Oleszikiewicz et al., 2017), für männliche Stimmen konnte kein Unterschied in Bezug auf die Wärme ausgemacht werden. Eine andere Studie zeigte, dass *vocal feminity* weniger kompetent bewertet wird als *vocal masculinity*, aber es konnte kein Unterschied hinsichtlich der Wärmebeurteilung ausgemacht werden (Ko et al., 2009). Krahé und Papakonstantinou (2020) zeigten hingegen, dass sich hohe Stimmen und tiefe Stimmen zwar hinsichtlich der Wärmebeurteilung, allerdings nicht bei der Kompetenzbeurteilung unterschieden. Die Befunde müssten künftig systematisch betrachtet werden. Nichtsdestotrotz können die Befunde in der Summe als Indizien dafür gedeutet werden, Kompetenz durch eine tiefe Stimme zu vermitteln sowie Wärme durch eine hohe Stimme. Ein Training zur gezielten Stimmmodulation kann deswegen im Lehrberuf (nicht nur zur Schonung der Stimme) sinnvoll sein.

> Die Sprechrate hat einen starken Effekt darauf, wie kompetent eine Stimme erscheint. Ein schneller Sprecher, der z. B. keine Füllsel wie „äh“, „ähm, „hmm“, „jaa“ verwendet, erscheint überzeugender und intelligenter. Wenige Sprechpausen und weniger Wiederholungen sowie eine dynamische Stimme signalisieren Kompetenz. Laute und männliche [tiefe] Stimmen werden – im Gegensatz zu weichen, sanften, weiblichen [hohen] Stimmen – als „konsequent wirkend“ wahrgenommen. [...] Intonation und Sprechrate sind aber stets durch die jeweilige Sprache, die Situation [...] und/oder durch den/die Zuhörer bestimmt, ähnlich der Lautstärke eines Sprechers, die themengemäß und raumangemessen sein muss. (Kiese-Himmel, 2016, S. 32)

Weitergehend kann auch die Kleidung Informationen zur Kompetenz einer Person liefern. Formale Kleidung kann (als Lehrperson) mit höherer Kompetenzzuschreibung einhergehen (Morris et al., 1996; Gorham et al., 1999), jedoch wird auch das Ausbleiben von Vorteilen formaler Kleidung beschrieben (Gorham et al., 1997). Gorham et al. (1999) erklären, dass der Zusammenhang von Kompetenzzuschreibung und formaler Kleidung durch das Alter der Dozierenden erklärt werden kann: Je älter die Dozierenden sind, desto formaler kleiden sie sich, und ältere Dozierende werden allgemein kompetenter beschrieben als jüngere. In Studie 4 konnte jedoch kein bedeutsamer Zusammenhang von Kompetenzzuschreibung

und Alter der Dozierenden ausgemacht werden, sodass angenommen werden kann, dass Kleidung als eine Moderatorvariable wirkt. Neben der formellen Kleidung kann die Farbe Rot offenbar zur Wahrnehmung von mehr Kompetenz und Kompetitivität führen (Maier et al., 2013; vgl. Elliot & Maier, 2014). Diese Hinweise müssten jedoch gezielt für den Lehr-Lern-Kontext untersucht werden, z. B. indem in experimentellen Untersuchungen Lehrpersonen systematisch andere Kleidung tragen.

Carli et al. (1995) verglichen unterschiedliche Referierendenstile miteinander. Dabei zeigte sich, dass Referierende mit einem „unterwürfigen“ Stil (*submissive style*) die geringsten Kompetenzzuschreibungen erhielten. Dieser Stil ist geprägt durch „a soft, pleading voice with many hesitations and stumbles, a slumped posture, nervous hand gestures, and an avert gaze“ (Carli et al., 1995, S. 1033). Die besten Bewertungen fielen auf den sozialen (*social style*) und den aufgabenorientierten Stil (*task style*):

> The social style included a voice of moderate volume, a relaxed posture with the body leaning toward the listener, a friendly facial expression, and a moderately high amount of eye contact. The task style included a rapid rate of speech, a firm tone of voice, moderate voice volume, few hesitations, an upright posture, calm hand gestures, and a moderately high amount of eye contact. (Carli et al., 1995, S. 1034)

Der Ausdruck von Allgegenwärtigkeit und Präsenz durch Lehrpersonen, der mit der Kompetenzwahrnehmung aufseiten der Lernenden einhergehen kann, sollte schon früh in die Ausbildung von Lehrpersonen eingehen, z. B. in der Vorbereitung auf erste Praxiserfahrungen im Lehramtsstudium (vgl. Hannemann, 2020). Eine zugewandte, offene Haltung mit moderater Körperspannung, eine der Situation angepasste Stimmmodulation sowie angemessene Kleidung können den ersten Eindruck, für den es sprichwörtlich keine zweite Chance gibt, positiv beeinflussen.

Auch wenn die bisherigen Studienergebnisse auf Ansätze zur Steigerung von Lernendenkompetenzerwartungen an Lehrpersonen hinweisen, müssen diese weiter untersucht werden. Welche Aspekte neben Stimme, Haltung, Positionierung und Kleidung relevant sind, müsste systematischer analysiert werden. Dafür ist es möglich, verschiedene Einflüsse auf die Kompetenzwahrnehmung offen zu erfragen (z. B. „Woran erkennen Sie beim Kennenlernen eine kompetente Lehrperson?“), die dann inhaltsanalytisch ausgewertet und weitergehend faktoranalytisch geclustert werden. So können verschiedene Kompetenzmarker untersucht und deren Bedeutung auch für unterschiedliche Lehr-Lern-Kontexte (z. B. Schule, Hochschule, Erwachsenenbildung) verglichen werden.

19.2 Lernendenkompetenzerwartungen in weiteren Ansätzen zur Gestaltung der Lehr-Lern-Interaktion

Neben den eher oberflächlichen Methoden, die auf den ersten Eindruck abzielen, ist es zudem wichtig, eine lernförderliche Lehr-Lern-Interaktion herzustellen. Dies ist sowohl Aufgabe der Lehrpersonen als auch der Lernenden. Die Bedeutung von Lernendenkompetenzerwartungen kann in verschiedenen Theorien wiedergefunden werden, auch wenn sie keine explizite Erwähnung erfahren. Folgend sollen beispielhaft Ansätze erläutert werden, bevor auf synergetisch-systemische Betrachtungen detaillierter eingegangen wird, um die oben genannten Hinweise zur Gestaltung des ersten Eindrucks theoretisch zu verorten und weitere Forschungsansätze aufzuzeigen, die das kognitiv geprägte Modell von Nickel (1985) ergänzen.

Das COACTIV-Modell (Baumert & Kunter, 2006; Kunter et al., 2011) listet und strukturiert verschiedene Aspekte der professionellen Kompetenz von Lehrpersonen. Neben dem als zentral angesehenen Professionswissen werden die motivationalen Orientierungen, die Überzeugungen und Werthaltungen sowie die selbstregulatorischen Fähigkeiten benannt. Zum Professionswissen gehört auch das pädagogisch-psychologische Wissen (Kompetenzbereich). Zu diesem Kompetenzbereich (in den Kompetenzfacetten „effektive Klassenführung“ und „Wissen über Lernprozesse“) können die hier erzielten Ergebnisse zugeordnet werden. Lehrpersonen sollte klar sein, dass nicht nur die Lehrendenerwartungen eine Rolle in der Interaktion spielen, sondern auch die Lernendenerwartungen, und dass Letztere sich u. a. auf den Lernerfolg der Lernenden auswirken können (Wissen über Lernprozesse; Zusammenhang von Lernendenkompetenzerwartungen und Lernerfolg bei Adediwura & Tayo, 2007; Edwards et al., 2009; Feldman & Prohaska, 1979; Jamieson et al., 1987). Die besondere Bedeutung der Wärmewahrnehmung für das sozial-emotionale Erleben in Lehr-Lern-Kontexten (vgl. Studie 4) ist auch ein Aspekt, der zum Klassenführungswissen gezählt werden kann. Welche weiteren Formen von Lernendenerwartungen, u. a. Kompetenzerwartungs-Erwartungen (s. Kapitel 4), bedeutsam sind, müsste zukünftig in weiteren Studien untersucht werden.

Marzano et al. (2003) beschreiben die Lehr-Lern-Beziehung als Schlüssel, um effektive Klassenführung umzusetzen. Der durchschnittliche Effekt der Lehr-Lern-Beziehung auf störungsfreien Unterricht liegt bei $d = 0.869$, doch je jünger die Schüler*innen sind, desto höher fallen die Effektstärken aus (von $d = 0.549$ bei High School bis $d = 2.891$ bei Middle School/Junior High). Um eine gute Lehr-Lern-Beziehung herzustellen, benennen Marzano et al. (2003) drei Handlungsschritte (*action steps*; vgl. Haag, 2018):

1. Etablierung eines angemessenen Grades an Dominanz
   - Festes, bestimmendes, durchsetzungsfähiges Verhalten ohne Aggressivität mit aufrechter Haltung, Blickkontakt sowie angemessenem Tonfall und klarer Aussprache
   - Klare, früh beschriebene Lernziele inklusive Feedback zur Zielerreichung
2. Kommunikation eines angemessenen Grades an Kooperation
   - Vorsehen flexibler Lernziele
   - Zeigen von Interesse an Lernenden
   - Präsenzmaßnahmen, z. B. Blickkontakt durch den ganzen Raum, sich im Raum bewegen
   - Angemessenes Reagieren auf inkorrekte Antworten (Positives hervorheben)
3. Berücksichtigung der Bedürfnisse von Lernendentypen und individuellen Besonderheiten der einzelnen Lernenden

Die Balance von Dominanz und Kooperation steht also nach Marzano et al. (2003) im Mittelpunkt für eine effektive Lehr-Lern-Interaktion. Dabei lassen sich die Maßnahmen, die mit der Kompetenz- und Wärmewahrnehmung im Zusammenhang stehen wie Körperhaltung und Stimmmodulation, direkt wiederfinden. Für Mazano et al. (2003) ist es für die Gestaltung der Lehr-Lern-Interaktion vonseiten der Lehrpersonen wichtig, ein angemessenes Maß an Dominanz und Macht (*dominance*) zu demonstrieren, das Kompetenz ausstrahlen und somit hohe Kompetenzerwartungen bei den Lernenden induzieren kann, sowie bereit dafür zu sein, die Führung zu übernehmen. Lehrpersonen müssen demnach angemessen mit ihren Machtmitteln umgehen lernen (vgl. Misamer, 2019).

Auch wenn in der Lehr-Lern-Situation die Lehrpersonen mehr Machtmittel zur Verfügung haben (vgl. Misamer, 2019; Steins, 2013; Thies, 2010), können Lernende das Lehr-Angebot mit den ihnen gegebenen Machtmitteln aktiv transformieren. Steins (2013) beschreibt unterschiedliche Formen von Machtmitteln, die auch Lernende haben. Es gibt für Lernende die Möglichkeit, (1.) durch soziale Anerkennung, Aufmerksamkeit und positives Feedback die Lehrperson zu belohnen (Belohnungsmacht), aber auch (2.) durch Ausbleiben der Mitarbeit, Sabotieren des Unterrichts sowie durch Gerüchte und Beleidigungen zu bestrafen (Bestrafungsmacht bzw. Macht zu zwingen). (3.) Weitergehend haben Schüler*innen in Schulen einen Rechtsanspruch auf Bildung, aber auch im Studium und in der Erwachsenenbildung kommt (mehr oder weniger) ein Vertrag zustande, der von beiden Seiten erfüllt werden muss (legitimierte Macht). Weitere Formen der Machtmittel sind (4.) die Referenzmacht (Beliebtheit, überzeugendes Auftreten), (5.) die Expertisemacht (Fähigkeiten und Kompetenz), (6.) die

Informationsmacht (Position, etwas mitteilen zu können) und (7.) die ökologische Macht (Fähigkeit zur Gestaltung der Umwelt). Je älter die Lernenden werden, desto stärker können sie mit ihren Machtmitteln wirken. Insbesondere die Expertise- und die Informationsmacht sind für ältere Lernende von besonderer Bedeutung in der Lehr-Lern-Interaktion sowie bei der Einflussnahme auf den Lehr-Lern-Prozess. Bei einer wahrgenommenen geringeren Fachkompetenz der Lehrpersonen, also geringen Lernendenkompetenzerwartungen, kann die eigene Expertisemacht genutzt werden, um den Unterricht für alle zu bereichern und die fehlende Fachkompetenz der Lehrperson auszugleichen. Dies sollte in einer empathischen, wertschätzenden Form passieren. Darüber hinaus sollte den Lernenden bewusst sein, dass Lehrpersonen neben der Fachkompetenz zumeist auch pädagogische Kompetenz (z. B. zur Anregung zum selbstgesteuerten Lernen) mitbringen (vgl. Baumert & Kunter, 2006; Helmke, 2017; Kunter et al., 2011) sowie Lernende und Lehrpersonen gemeinsam ein positives, lernförderliches Klima aufbauen können. Sollten Lernende mit der fachlichen Expertise der Lehrperson in einer Lernsituation nicht zufrieden sein, haben sie meist die Möglichkeit, mit den Lehrpersonen über die Gestaltung der Lehre zu sprechen (insbesondere in der Erwachsenenbildung; Informationsmacht). Um dies erfolgreich umzusetzen, sind Empathie für die Lehrperson und Perspektivwechsel förderlich (vgl. Steins, 2013). So könnten im Hochschulkontext Studierende in Absprachen mit Dozierenden treten, wie die Lehr-Lern-Situation konkret umgesetzt wird, z. B. selbstgesteuertes Lernen versus Frontalunterricht, und welches Prüfungsformat unter Berücksichtigung der Modulhandbücher angemessen ist (im Sinne des *constructive alignment*; vgl. Biggs & Tang, 2011). Es kann also argumentiert werden, dass Lehrpersonen ihre Kompetenzen transparent machen sowie professionell mit Schwächen und Unwissen umgehen sollten (vgl. Thies et al., 2016).

Ein weiteres Modell, in dem Lernendenkompetenzerwartungen relevant sind, ist das Modell von Helmke (2017) zur Wirkungsweise von Unterricht. Nach Helmke (2017) unterbreitet die Lehrperson mit ihrem Unterricht ein Angebot. Dieses Angebot muss jedoch wahrgenommen und zudem genutzt werden, was wiederum vom Lernpotenzial der lernenden Personen abhängt, zentral sind hierbei „Vorkenntnisse, Sprache(n), Intelligenz, Lern- und Gedächtnisstrategie; Lernmotivation, Anstrengungsbereitschaft, Ausdauer, Selbstvertrauen" (Helmke, 2017, S. 71). Hier müssten die kognitiven Schemata ergänzt werden – sowohl interaktional, relational (vgl. Thies, 2010) als auch normativ im Sinne von Rosemann (1978) sowie hinsichtlich der (Kompetenz- und Wärme-) Erwartungen (in Anlehnung an das *stereotype content model*; s. Kapitel 5). In dem heuristischen Modell von Aldrup (2017) wird davon ausgegangen, dass die Lehr-Lern-Interaktion und -Beziehung psychologische Grundbedürfnisse wie

Kompetenz- und Autonomieerleben sowie soziale Eingebundenheit (gemäß der Selbstbestimmungstheorie von Deci & Ryan, 2012) befriedigen kann und dadurch zu positiven Outcomes sowohl aufseiten der Lernenden als auch der Lehrpersonen (insbesondere deren berufliches Wohlbefinden) führt. Diese Outcomes wiederum beeinflussen die Lehr-Lern-Beziehung.

Das Ziel von Lehrpersonen und Lernenden muss es sein, gemeinsam Situationen zu schaffen, in denen sowohl Lernen als auch ein gutes Lehr-Lern-Klima ermöglicht werden. Lernende sollten ihre Schemata hinterfragen, z. B.: Muss die Lehrperson mir Wissen vermitteln oder die Lernsituation herstellen? Wie eigenständig muss ich mich selbst in den Lernprozess integrieren? Dürfen Lehrpersonen (auch fachliche) Schwächen haben?

19.3 Synergetisch-systemische Betrachtungen auf die Lehr-Lern-Interaktion

Das Vorgehen der durchgeführten Studien orientierte sich an deterministisch-kausalen Untersuchungen und Überlegungen, die experimentelle methodische Standards erfüllen (vgl. Döring & Bortz, 2016), und basierte auf der kognitiven Perspektive des transaktionalen Modells der Lehr-Lern-Interaktion von Nickel (1985). Alternativ kann für die gewählte Fragestellung eine übergreifende systemische Perspektive eingenommen (u. a. Hubrig & Herrmann, 2014) werden. Soziale Systeme – also auch die Lehr-Lern-Interaktion resp. die Lehr-Lern-Gruppe – sind dynamisch, komplex und nicht-linear (u. a. Strunk & Schiepek, 2006). So lassen sich Nickels (1985) Überlegungen auch in synergetisch-systemische Konzepte übersetzen (vgl. Thies, 2017).

Im Rahmen der Synergetik (Haken & Schiepek, 2010; Schiepek et al., 2013) lassen sich für die Lehr-Lern-Interaktion zwei Ebenen unterscheiden: (1.) eine mikroskopische Ebene, die durch psychische Systeme („Elemente“), die miteinander in Beziehung stehen, gekennzeichnet ist, und (2.) eine makroskopische Ebene, die durch spezifische Muster der Interaktion der Elemente untereinander bestimmt ist. Die Elemente innerhalb der Lehr-Lern-Interaktion sind die Lehrpersonen und Lernenden. Soziale Systeme bestehen aus mindestens zwei psychischen Systemen (bzw. Personen), die in gegenseitiger Wechselwirkung stehen („Zirkularität“). Psychische Systeme sind durch Kognitionen (darunter Wahrnehmung und Erwartungen), Emotionen und Verhalten gekennzeichnet, die als Elemente des psychischen Systems in zirkulärer Beziehung stehen und Kognitions-Emotions-Verhaltens-Muster (Tschacher, 1997) bilden. Die in der Studie 3 diskutierten unterschiedlichen „Langeweileprofile“ können als

unterschiedliche Formen der Kognitions-Emotions-Verhaltens-Muster betrachtet werden. Die zirkuläre Beziehung und wechselseitige Bezogenheit von Lernenden und Lehrpersonen sind Kern des Modells von Nickel (1985).

Die Interaktionsmuster zwischen Lehrpersonen und Lernenden sind selbstorganisiert: Die Lerngruppe inklusive der Lehrperson(en) schafft sich eigenständig (und häufig automatisiert) gemeinsame Wirklichkeitskonstruktionen, implizite und explizite Regeln sowie Routinen („Attraktoren" bzw. „Ordner"), die nicht von außen vorgegeben werden (vgl. Schallnus, 2006). Diese beobachtbaren Ordnungsmuster entstehen mit der Zeit, d. h., sie emergieren und reduzieren die Freiheitsgrade der Interaktion („Versklavungsprinzip"; Haken & Schiepek, 2010) und machen sie somit vorhersagbarer. Welche von vielen möglichen „Ordnern", also z. B. welche Routinen, sich in einem sozialen System durchsetzen, ist von vielen Faktoren abhängig. Auch Stereotype, u. a. Geschlechts-Professions-Assoziationen (vgl. Studie 1 und Studie 2), einzelner Personen können als intrasystemische „Kontrollparameter" fungieren und Interaktionsmuster emergieren lassen. Generell ist aufgrund der stärkeren Machtposition der Lehrpersonen (u. a. Misamer, 2019; Steins, 2013; Thies, 2010) im Vergleich zur Position der Lernenden die Wahrscheinlichkeit höher, dass intrasystemische Ordnungsbildung häufiger vonseiten der Lehrpersonen ausgeht. Dies kann ein Grund sein, warum die Lehrpersonen in Studie 4 weniger von den Lernendenkompetenzerwartungen beeinflusst sind. Da mehrere Ordner in Konkurrenz miteinander stehen, kann es sein, dass sich das Ordnungsmuster mit der Zeit oder je nach Situation verändert.

Wenngleich soziale Systeme selbstorganisiert sind, stehen sie dennoch im Austausch mit der Umwelt (Energie, Informationen, Ressourcen). Der soziokulturelle Bezugsrahmen nach Nickel (1985) findet in der synergetischen Theorie in der Umwelt des Systems seine Entsprechung. So können Informationen von außen – wie dies bei einer Erwartungsmanipulation im Sinne des Pygmalion-Effekts nach Rosenthal und Jacobsen (1966, 1971) der Fall war – auf das Lehr-Lern-System als Kontrollparameter wirken und weitergehend neue Ordnungszustände anregen. Auch in Studie 3 konnte durch die Kompetenzmanipulation als Kontrollparameter von außen das Verhalten der Teilnehmenden beeinflusst werden. Da soziale Systeme aber nicht-trivial und selbstreferenziell sind (von Foerster, 1997), d. h. immer in Abhängigkeit des aktuellen Systemzustands reagieren, kann keine gezielte Einflussnahme von außen erfolgen, sondern lediglich eine Verstörung („Perturbation"), die eine wie auch immer geartete Änderung des Ordnungszustands nach sich zieht. Dadurch lässt sich theoretisch erklären, warum Erwartungsmanipulationen zum Teil die Lehr-Lern-Interaktion beeinflussen können und zum Teil nicht. So zeigte sich die selbsterfüllende Prophezeiung bei Rosenthal und Jacobson (1966, 1971) lediglich in den Schulklassen 1 und 2, aber

nicht mehr in den Klassenstufen 3 bis 6. Dies kann auf Basis der Synergetik wie folgt erklärt werden: Im Laufe der Zeit verfestigen sich die Interaktionsmuster zwischen Lehrpersonen und Lernenden. Dabei wird das psychische System der Lehrpersonen so beeinflusst, dass sich ein stabiles Kognitions-Emotions-Verhaltens-Muster innerhalb der Lehr-Lern-Interaktion ausbildet und die bisher aufgebauten Gedanken über und Verhalten gegenüber den Lernenden durch die Kompetenzmanipulation nicht perturbiert werden kann. Der innere Systemzustand ist also nicht sensibel für den potenziellen Kontrollparameter aus der Umwelt.

Die widersprüchlichen Studienergebnisse von Feldman und Prohaska (1979) und Feldman und Theiss (1982) über das Erleben von Lehrpersonen können mithilfe der Synergetik ebenfalls erklärt werden. Bei Feldman und Prohaska (1979, Studie 2) wurde nicht direkt die Auswirkung der Lernendenkompetenzerwartungen auf die Lehrpersonen betrachtet. So waren Lehrpersonen entweder einem zugewandten oder einem nicht-zugewandten Verhalten von vermeintlichen Lernenden (und tatsächlichen Konföderierten) ausgesetzt. Es wurde zwar vorher festgestellt, dass sich Lernende, die eine kompetente Lehrperson erwarteten, zugewandter verhalten (Feldman & Prohaska, 1979, Studie 1), aber eine Kompetenzerwartung hielten die Konföderierten nicht. Das zugewandtere Verhalten kann als intrasystemischer Kontrollparameter das Kognitions-Emotions-Verhaltens-Muster der Lehrperson „versklavt" haben, sodass sie sich wohlfühlte, die lernende Person sympathischer fand und sich adäquater verhielt. Feldman und Theiss (1982) wiederum manipulierten sowohl die Kompetenzerwartung der Lernenden als auch der Lehrpersonen. Die Kompetenzmanipulation für die Lehrpersonen konnte als äußerer Kontrollparameter das Erleben der Lehrpersonen „ordnen", während die Kompetenzmanipulation der Lernenden das Erleben der Lernenden selbst anregte. Durch die stärkere Machtposition der Lehrperson konnte sich der Pygmalion-Effekt auf Basis der Lehrendenkompetenzerwartungen durchsetzen, während die Manipulation der Lernendenkompetenzerwartungen in der kurzen Zeit „nur" das psychische System der Lernenden beeinflusste und keinen Ordnungszustand des sozialen Systems ermöglichen konnte.

In der Studie 4 wurden universitäre Lehrveranstaltungen nach mindestens einer, immer aber innerhalb der ersten fünf Sitzungen untersucht, in denen die Lernenden die jeweilige Lehrperson schon kannten (im Mittel bereits seit etwa einem halben Jahr). In dieser Zeit konnten erste Lernendenkompetenzerwartungen an Lehrpersonen die Lehr-Lern-Interaktion bereits beeinflussen und es zeigten sich kleine bis mittelgroße Effekte auf das Erleben positiver Emotionen, von Flow, von Störungen und Störungsmissmanagement sowie auf die eigene Kompetenzwahrnehmung aufseiten der Lehrpersonen. Die Lernendenkompetenzerwartungen auf der Mikroebene konnten somit theoretisch

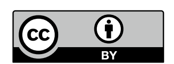
als systemimmanenter Kontrollparameter (vgl. Haken & Schiepek, 2010) die Makroebene, also die Lehr-Lern-Interaktionsmuster, „mitordnen“. Im Vergleich zu Feldman und Prohaska (1979) sowie Feldman und Theiss (1982) konnten die Lernendenkompetenzerwartungen in der Lehr-Lern-Interaktion zeitlich länger wirksam sein und stabilere Muster herausbilden. Welche Rolle die Zeit bei der Wirkung von Lernendenkompetenzerwartungen spielt, muss in weiteren Studien geklärt werden. Konkret braucht es also wie bereits erwähnt längsschnittliche Untersuchungen, um die Wirkung der Kompetenzerwartungen bei längerfristigen Lehr-Lern-Interaktionen wie in Schule, Hochschule und Erwachsenenbildung ausmachen zu können.

Die Prozesshaftigkeit der Lehr-Lern-Interaktion kann außerdem z. B. mit *state space grids* analysiert werden (Pennings & Mainhard, 2016). Mit *state space grids* können Zustände eines Systems und Abfolgen der Zustände grafisch in einem zweidimensionalen Gitter dargestellt werden. So ist es denkbar, das Verhalten von Lehrpersonen und Lernenden zu betrachten, wenn die Kompetenzerwartungen der Lernenden manipuliert wurden. Auch denkbar wären Eyetracking-Studien, um den Blickkontakt in Abhängigkeit der Kompetenzerwartungen in Lehr-Lern-Interaktionen genauer zu untersuchen. Durch diese Formen der Untersuchungen könnte ausgemacht werden, welche Reaktionen der Lehrpersonen negative Interaktionsspiralen (vgl. Nickel, 1985) vorbeugen bzw. positive aufbauen. Nichtsdestotrotz ist es für empirische Überprüfungen unmöglich, alle Faktoren eines sozialen Systems auszumachen oder gar zu kontrollieren. Deshalb es ist sinnvoll, einen zirkulären Prozess in seine Teile zu gliedern, um systematisch relevante Teilbereiche zu identifizieren und somit ein System beschreibbar zu machen (Kriz, 2007). So schreibt auch Witte (2017, S. 1): „Dieser qualitative, nicht überbrückbare Unterschied zwischen einer elementaristisch-kausalen und einer synergetisch-systemischen Position existiert nicht. Es lässt sich vielmehr die Haltung rechtfertigen, dass die elementaristisch-kausale Position eine Vereinfachung der synergetisch-systemischen Betrachtung darstellt.“

19.4 Resümierendes Fazit

Die Untersuchungen zu Erwartungseffekten in der Lehr-Lern-Interaktion haben sich seit Rosenthal und Jacobson (1966, 1971) in mehrere Richtungen entwickelt: In einigen wissenschaftlichen Betrachtungen stehen die Lehrendenerwartungen im Fokus, in anderen die Lernendenerwartungen (vgl. u. a. Ludwig, 2018). Lediglich in zwei Studien wurden die Lernendenkompetenzerwartungen als Prädiktor für Lehrendenerleben und -verhalten betrachtet (Feldman & Prohaska,

1979; Feldman & Theiss, 1982). Diese Forschungslücke wurde auf Basis des transaktionalen Modells der Lehr-Lern-Interaktion von Nickel (1985) mit der vorliegenden Arbeit mit aufeinander aufbauenden Studien adressiert. Lernendenkompetenzerwartungen haben einen Einfluss auf die Interaktionsgestaltungen im Lehr-Lern-Kontext, jedoch muss auf Basis der erzielten Daten von einem kleinen Effekt auf das Erleben und Verhalten von Lehrpersonen ausgegangen werden. Inwieweit sich dieser Effekt langfristig verkleinert, verstetigt oder sogar vergrößert, kann auf Basis der durchgeführten Studien und vorangehender Untersuchungen nicht vorhergesagt werden. Weitergehend bleibt die Frage offen, wie andere Erwartungen, z. B. hinsichtlich Wärme und Vertrauenswürdigkeit, in Interaktion mit den Kompetenzerwartungen aufseiten der Lernenden bedeutsam sind. Außerdem gibt es weitere Erwartungen, die in der Lehr-Lern-Interaktion relevant sein könnten, wie z. B. Erwartungs-Erwartungen. Es braucht mehr Forschung, um die Bedeutung der Lernendenerwartungen für die Lehr-Lern-Interaktion beschreibbar zu machen.



Literatur

Adediwura, A. A., & Tayo, B. (2007). Perception of teachers' knowledge, attitude and teachings skills as predictor of academic performance in Nigerian secondary schools. *Educational Research and Review, 2*(7), 165–171. https://doi.org/10.5897/ERR.9000250

Aldrup, K. (2017). *The teacher-student relationship is beneficial for both sides: Associations with students' school adjustment and teachers' occupational well-being* [Dissertation, Christian-Albrechts-Universität zu Kiel]. MACAU Publikationsserver der Christian-Albrechts-Universität zu Kiel. https://nbn-resolving.org/urn:nbn:de:gbv:8-diss-220671

Anderson, K. J., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences, 27*(2), 184–201. https://doi.org/10.1177%2F0739986304273707

Apel, H., & Sandfuchs, U. (2003). Guter Unterricht. Sichtweisen, Befunde, Hinweise. *Lernchancen, 31*, 4–13.

Argyle, M. (1972). *Soziale Interaktion.* Kiepenheuer & Witsch.

Asbrock, F. (2010). Stereotypes of social groups in Germany in terms of warmth and competence. *Social Psychology, 41*(2), 76–81. https://doi.org/10.1027/1864-9335/a000011

Autorengruppe Fachkräftebarometer (2019). *Fachkräftebarometer Frühe Bildung 2019*. Weiterbildungsinitiative Frühpädagogische Fachkräfte. https://www.weiterbildungsinitiative.de/publikationen/detail/fachkraeftebarometer-fruehe-bildung-2019

Auwarter, A. E., & Aruguete, M. S. (2008). Effects of student gender and socioeconomic status on teacher perceptions. *Journal of Educational Research, 101*(4), 242–246. https://doi.org/10.3200/JOER.101.4.243-246

Babad, E. Y., Inbar, J., & Rosenthal, R. (1982). Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers. *Journal of Educational Psychology, 74*(4), 459–474. https://doi.org/10.1037/0022-0663.74.4.459

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2016). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (14. Aufl.). Gabler. https://doi.org/10.1007/978-3-662-46076-4

Balliet, D., Wu, J., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin, 140*(6), 1556–1581. https://doi.org/10.1037/a0037737

Bandura, A. (1977a). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1977b). *Social learning theory*. Prentice Hall.

M. Hackbart, *Die Lehrperson als Golem oder Galatea*,
https://doi.org/10.1007/978-3-658-39444-8

Barmaki, R. (2014). Nonverbal communication and teaching performance. *Proceeding of the 7th International Conference on Educational Data Mining* (EDM), 441–443. https://educationaldatamining.org/EDM2014/uploads/procs2014/YRT/441_EDM-2014-Full-Proceedings.pdf

Basow, S. A., Codos, S., & Martin, J. L. (2013). The effects of professors' race and gender on student evaluations and performance. *College Student Journal, 47*(2), 352–363.

Basow, S. A., Phelan, J. E., & Capotosto, L. (2006). Gender patterns in college students' choices of their best and worst professors. *Psychology of Women Quarterly, 30*(1), 25–35. https://doi.org/10.1111%2Fj.1471-6402.2006.00259.x

Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology, 79*(3), 308–314. https://doi.org/10.1037/0022-0663.79.3.308

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 9*(4), 469–520. https://doi.org/10.1007/s11618-006-0165-2

Bavishi, A., Madera, J. M., & Hebl, M. R. (2010). The effect of professor ethnicity and gender on student evaluations: Judged before met. *Journal of Diversity in Higher Education, 3*(4), 245–256. https://doi.org/10.1037/a0020763

Beelmann, A., & Neudecker, C. (2020). Entwicklungspsychologische Grundlagen für die Entstehung von Vorurteilen. In L.-E. Petersen & B. Six (Hrsg.), *Stereotype, Vorurteile und soziale Diskriminierung. Theorien, Befunde und Interventionen* (2. Aufl., S. 113–124). Beltz.

Beishuizen, J. J., Hof, E., van Putten, C. M., Bouwmeester, S., & Asscher, J. J. (2001). Students' and teachers' cognitions about good teachers. *British Journal of Educational Psychology, 71*(2), 185–201. https://doi.org/10.1348/000709901158451

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J*ournal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology, 74*(2), 170–179. https://doi.org/10.1037/0022-0663.74.2.170

Berthold, C., Kessler, M. S., Kreft, A.-K., & Leichsenring, H. (2011). *Schwarzer Peter mit zwei Unbekannten. Ein empirischer Vergleich der unterschiedlichen Perspektiven von Studierenden und Lehrenden auf das Studium.* Centrum für Hochschulentwicklung. https://www.che.de/download/che_ap141__doppelbefragung-pdf/

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology, 60*(4), 485–499. https://doi.org/10.1037/0022-3514.60.4.485

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university. What the student does* (4. Aufl.). Open University Press.

Blatter, K., & Cajochen, C. (2007). Circadian rhythms in cognitive performance: Methodological constraints, protocols, theoretical underpinnings. *Physiology & Behavior, 90*(2–3), 196–208. https://doi.org/10.1016/j.physbeh.2006.09.009

Bonds-Raacke, J., & Raacke, J. D. (2007). The relationship between physical attractiveness of professors and students' ratings of professor quality. *Journal of Psychiatry, Psychology and Mental Health, 1*(2), 1–7.

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Springer. https://doi.org/10.1007/978-3-642-12770-0

Bosak, J., Sczesny, S., & Eagly, A. H. (2008). Communion and agency judgments of women and men as a function of role information and response format. *European Journal of Social Psychology, 38*(7), 1148–1155. https://doi.org/10.1002/ejsp.538

Brophy, J. E., & Good, T. L. (1976). *Die Lehrer-Schüler-Interaktion.* Urban & Schwarzenberg.

Brunner, E. J., & Huber, G. L. (1991). Störungen der Schulklasse als soziales System: Intervention. In M. Perrez & U. Baumann (Hrsg.), *Klinische Psychologie. Intervention* (Band 2, S. 386–394). Huber.

Budde, J. (2008). *Bildungs(miss)erfolge von Jungen und Berufswahlverhalten bei Jungen/männlichen Jugendlichen.* Bundesministerium für Bildung und Forschung. https://www.bildungsserver.de/onlineressource.html?onlineressourcen_id=39841

Bühner, M. (2010). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). Pearson.

Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. Pearson.

Bundesagentur für Arbeit (2021). *Die Arbeitsmarksituation von Frauen und Männern 2020. Nürnberg: Bundesagentur für Arbeit*. Bundesagentur für Arbeit. https://statistik.arbeitsagentur.de/DE/Statischer-Content/Statistiken/Themen-im-Fokus/Frauen-und-Maenner/generische-Publikationen/Frauen-Maenner-Arbeitsmarkt.pdf?__blob=publicationFile&v=10

Burke, K. (2012). *Big Five personality traits and astrology: The relationship between the moon variable and the NEO PI-R* (Publikationssnummer 3534069) [Dissertation, Pacifica Graduate Institute]. ProQuest Dissertation & Theses Open.

Bye, H. H., Herrebrøden, H., Hjetland, G. J., Røyset, G. Ø., & Westby, L. L. (2014). Stereotypes of Norwegian social groups. *Scandinavian Journal of Psychology, 55*(5), 469–476. https://doi.org/10.1111/sjop.12141

Carlsson, R. (2013). *Warmth and competence in implicit stereotypes and discrimination* [Thesis, Lund University]. LUP Lund University Publications. https://lup.lub.lu.se/record/3410013

Carli, L. L., LaFleur, S. J., & Loeber, C. C. (1995). Nonverbal behavior, gender, and influence. *Journal of Personality and Social Psychology, 68*(6), 1030–1041. https://doi.org/10.1037/0022-3514.68.6.1030

Casero, A. (2010). Modulating factors in the perception of teaching. *RELIEVE, 16*(2), Artikel 3. https://doi.org/10.7203/relieve.16.2.4135

Chaikin, A. L., & Derlega, V. J. (1974). Variables affecting the appropriateness of self-disclosure. *Journal of Consulting and Clinical Psychology, 42*(4), 588–593. https://doi.org/10.1037/h0036614

Chen, X., Zou, K., Abdullah, N., Whiteside, N., Sarmanova, A., Dhoerty, M., & Zhang, W. (2017). The placebo effect and its determinants in fibromyalgia: Meta-analysis of randomised controlled trials. *Clinical Rheumatology, 36*(7), 1623–1630. https://doi.org/10.1007/s10067-017-3595-8

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Erlbaum.

Cohrs, J. C., & Asbrock, F. (2009). Right-wing authoritarianism, social dominance orientation and prejudice against threatening and competitive ethnic groups. *European Journal of Social Psychology, 39*(2), 270–289. https://doi.org/10.1002/ejsp.545

Colagiuri, B., Livesey, E. J., & Harris, J. A. (2011). Can expectancies produce placebo effects for implicit learning? *Psychonomic Bulletin & Review, 18*(2), 399–405. https://doi.org/10.3758/s13423-010-0041-1

Cook, K. F., Kallen, M. A., & Amtmann, D. (2009). Having a fit: Impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Quality of Life Research, 18*(4), 447–460. https://doi.org/10.1007/s11136-009-9464-4

Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research, 77*(1), 113–143. https://doi.org/10.3102%2F003465430298563

Cragg, J. J., Warner, F. M., Finnerup, N. B., Jensen, M. P., Mercier, C., Richards, J. S., Wrigley, P., Soler, D., & Kramer, J. L. K. (2016). Meta-analysis of placebo responses in central neuropathic pain. Impact of subject, study, and pain characteristics. *PAIN, 157*(3), 530–540. https://doi.org/10.1097/j.pain.0000000000000431

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*(4), 631–648. https://doi.org/10.1037/0022-3514.92.4.631

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the bias map. *Advances in Experimental Social Psychology, 40*, 61–149. https://doi.org/10.1016/S0065-2601(07)00002-0

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., Bond, M. H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T. T., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., … Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology, 48*(1), 1–33. https://doi.org/10.1348/014466608X314935

Deci, E. L., & Ryan, R. M. (2012). *Self-determination theory.* In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Hrsg.), *Handbook of theories of social psychology* (Band 1, S. 416–436). Sage Publications. https://doi.org/10.4135/9781446249215.n21

Degner, J., & Dalege, J. (2013). The apple does not fall far from the tree, or does it? A meta-analysis of parent-child similarity in intergroup attitudes. *Psychological Bulletin, 130*(6), 1270–1304. https://doi.org/10.1037/a0031436

Degner, J., Mangels, J., & Zander, L. (2019). Visualizing gendered representations of male and female teachers using a reverse correlation paradigm. *Social Psychology, 50*(4), 233–251. https://doi.org/10.1027/1864-9335/a000382

Denner, S., Peinelt, N., & Sacher, W. (2003). Unterricht aus Schülersicht. Erste Ergebnisse aus einem laufenden Forschungsprojekt. *Lernchancen, 31*, 50–53.

DiDonato, T. E., Ullrich, J., & Krueger, J. I. (2011). Social perception as induction and inference: An integrative model of intergroup differentiation, ingroup favoritism, and differential accuracy. *Journal of Personality and Social Psychology, 100*(1), 66–83. https://doi.org/10.1037/a0021051

Diefenbach, H. (2010). Jungen – die „neuen" Bildungsverlierer. In G. Quenzel & K. Hurrelmann (Hrsg.), *Bildungsverlierer. Neue Ungleichheiten* (S. 245–271). VS Verlag. https://doi.org/10.1007/978-3-531-92576-9_12

Diehl, M., Owen, S. K., & Youngblade, L. M. (2004). Agency and communion attributes in adults' spontaneous self-representations. *International Journal of Behavioral Development, 28*(1), 1–15. https://doi.org/10.1080%2F01650250344000226

Diewald, G., & Steinhauer, A. (2020). *Handbuch geschlechtergerechte Sprache. Wie Sie angemessen und verständlich gendern.* Dudenverlag

Dimitriadis, P. A., & Panagiotis, Z. (2017). Nocebo effect in Menière's disease: A meta-analysis of placebo-controlled randomized controlled trials. *Otology & Neurotology, 38*(9), 1370–1375. https://doi.org/10.1097/MAO.0000000000001555

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285–290. https://doi.org/10.1037/h0033731

Ditton, H. (2002). Lehrkräfte und Unterricht aus Schülersicht – Ergebnisse einer Untersuchung im Fach Mathematik. *Zeitschrift für Pädagogik, 48*(2), 262–286. https://doi.org/10.25656/01:3833

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Springer. https://doi.org/10.1007/978-3-642-41089-5

Duden (o. J.). *Erwartung, die.* Abgerufen am 01.02.2020 von https://www.duden.de/rechtschreibung/Erwartung

Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J. C., Akande, A. D., Adetoun, B. E., Adewuyi, M. F., Tserere, M. M., Al Ramiah, A., Mastor, K. A., Berlow, F. K., Bonn, G., Tafarodi, R. W., Bosak, J., Cairns, E., Doherty, C., Capozza, D., Chandran, A., Chryssochoou, X., … Storari, C. C. (2013). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap. *British Journal of Social Psychology, 52*(4), 726–746. https://doi.org/10.1111/bjso.12005

Dusek, J. B., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology, 75*(3), 327–346. https://doi.org/10.1037/0022-0663.75.3.327

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110*(1), 109–128. https://doi.org/10.1037/0033-2909.110.1.109

Eagly, A. H., & Mladinic, A. (1994). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *European Review of Social Psychology, 5*(1), 1–35. https://doi.org/10.1080/14792779543000002

Eckes, T. (2002). Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model. *Sex Roles, 47*(3), 99–114. https://doi.org/10.1023/A:1021020920715

Eckes, T. (2020). Messung von Stereotypen. In L.-E. Petersen & B. Six (Hrsg.), *Stereotype, Vorurteile und soziale Diskriminierung. Theorien, Befunde und Interventionen* (2. Aufl., S. 99–108). Beltz.

Eden, D., & Zuk, Y. (1995). Seasickness as a self-fulfilling prophecy: Raising self-efficacy to boost performance at sea. *Journal of Applied Psychology, 80*(5), 628–635. https://doi.org/10.1037/0021-9010.80.5.628

Edwards, A., Edwards, C., Shaver, C., & Oaks, M. (2009). Computer-mediated word-of-mouth communication on ratemyprofessors.com: Expectancy effects on student cognitive and behavioral learning. *Journal of Computer-Mediated Communication, 14*(2), 368–392. https://doi.org/10.1111/j.1083-6101.2009.01445.x

Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden* (5. Aufl.). Beltz.

Elashoff, J. D., & Snow, R. E. (1972). *Pygmalion auf dem Prüfstand*. Kösel.

Elliot, A. J., & Maier, M. A. (2014). Color psychology: Effects of perceiving color on psychological functioning in humans. *Annual Review of Psychology, 65*, 95–120. https://doi.org/10.1146/annurev-psych-010213-115035

Feldman, R. S., & Prohaska, T. (1979). The student as Pygmalion: Effects of student expectation on the teacher. *Journal of Educational Psychology, 71*(4), 485–493. https://doi.org/10.1037/0022-0663.71.4.485

Feldman, R. S., Saletsky, R. D., Sullivan, J., & Theiss, A. (1983). Student locus of control and response to expectations about self and teacher. *Journal of Educational Psychology, 75*(1), 27–32. https://doi.org/10.1037/0022-0663.75.1.27

Feldman, R. S., & Theiss, A. J. (1982). The teacher and student as Pygmalions: Joint effects of teacher and student expectations. *Journal of Educational Psychology, 74*(2), 217–223. https://doi.org/10.1037/0022-0663.74.2.217

Fend, H. (2009). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen* (2. Aufl.). VS Verlag. https://doi.org/10.1007/978-3-531-91788-7

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. Zanna (Hrsg.), *Advances in experimental social psychology* (Band 23, S. 1–74). Academic Press. https://doi.org/10.1016/S0065-2601(08)60317-2

Fiske, S. T., & Taylor, S. E. (2017). *Social cognition. From brains to culture* (3. Aufl.). Sage Publications.

Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues, 55*(3), 473–489. https://doi.org/10.1111/0022-4537.00128

Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology, 26*, 21–42. https://doi.org/10.1146/annurev.soc.26.1.21

Friedrich, M. C. G. (2017). *Textverständlichkeit und ihre Messung*. Waxmann.

Fries, S., Horz, H., & Haimerl, C. (2006). Pygmalion in media-based learning: Effects of quality expectancies on learning outcomes. *Learning and Instruction, 16*(4), 339–349. https://doi.org/10.1016/j.learninstruc.2006.07.005

Frisch, M. (1999). *Andorra*. Suhrkamp.

Froehlich, L., Martiny, S. E., Deaux, K., & Mok, S. Y. (2016). "It's their responsibility, not ours". Stereotypes about competence and causal attributions for immigrants' academic underperformance. *Social Psychology, 47*(2), 74–86. https://doi.org/10.1027/1864-9335/a000260

Froming, W. J., Nasby, W., & McManus, J. (1998). Prosocial self-schemas, self-awareness, and children's prosocial behavior. *Journal of Personality and Social Psychology, 75*(3), 766–777. https://doi.org/10.1037/0022-3514.75.3.766

Geis, F. L. (1993). Self-fulfilling prophecies: A social psychological view of gender. In A. E. Beall & R. J. Sternberg (Hrsg.), *The psychology of gender* (S. 9–54). Guilford Press.

Geiser, C. (2010). *Datenanalyse mit MPlus. Eine anwendungsorientierte Einführung*. VS Verlag. https://doi.org/10.1007/978-3-531-92042-9

Gentrup, S., Rjosk, C., Stanat, P., & Lorenz, G. (2018). Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen. *Zeitschrift für Erziehungswissenschaft, 21*(4), 867–891. https://doi.org/10.1007/s11618-018-0806-2

Gilovich, T., Keltner, D., & Nisbett, R. E. (2006). *Social Psychology*. Norton.

Glascock, J., & Ruggiero, T. E. (2006). The relationship of ethnicity and sex to professor credibility at a culturally diverse university. *Communication Education, 55*(2), 197–207. https://doi.org/10.1080/03634520600566165

Glock, S. (2016). Does ethnicity matter? The impact of stereotypical expectations on in-service teachers' judgments of students. *Social Psychology of Education, 19*(3), 493–509. https://doi.org/10.1007/s11218-016-9349-7

Glock, S., & Kleen, H. (2019). Attitudes toward students from ethnicity minority groups: The roles of preservice teachers' own ethnic backgrounds and teacher efficacy activation. *Studies in Educational Evaluation, 62*, 82–91. https://doi.org/10.1016/j.stueduc.2019.04.010

Glock, S., & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations on student techers' judgements. *Social Psychology of Education, 16*(1), 111–127. https://doi.org/10.1007/s11218-012-9197-z

Glock, S., & Krolak-Schwerdt, S. (2014). Stereotype activation versus application: How teachers process and judge information about students from ethnic minorities and with low socioeconomic background. *Social Psychology of Education, 17*(4), 589–607. https://doi.org/10.1007/s11218-014-9266-6

Goebel, B. L., & Cashen, V. M. (1979). Age, sex, and attractiveness as factors in student ratings of teachers: A developmental study. *Journal of Educational Psychology, 71*(5), 646–653. https://doi.org/10.1037/0022-0663.71.5.646

Gollwitzer, A., & Bargh, J. A. (2018). Social psychological skill and its correlates. *Social Psychology, 49*(2), 88–102. https://doi.org/10.1027/1864-9335/a000332

Gorham, J., Cohen, S. H., & Morris, T. L. (1997). Fashion in the classroom II: Instructor immediacy and attire. *Communication Research Reports, 14*(1), 11–23. https://doi.org/10.1080/08824099709388641

Gorham, J., Cohen, S. H., & Morris, T. L. (1999). Fashion in the classroom III: Effects of instructor attire and immediacy in natural classroom interactions. *Communication Quarterly, 47*(3), 281–299. https://doi.org/10.1080/01463379909385560

Greitemeyer, T. (2012). *Sozialpsychologie*. Kohlhammer.

Greitemeyer, T. (2020). Sich selbst erfüllende Prophezeiungen. In L.-E. Petersen & B. Six (Hrsg.), *Stereotype, Vorurteile und soziale Diskriminierung. Theorien, Befunde und Interventionen* (2. Aufl., S. 82–89). Beltz.

Griffin, B. W. (2001). Instructor reputation and student ratings of instruction. *Contemporary Educational Psychology, 26*(4), 534–553. https://doi.org/10.1006/ceps.2000.1075

Grouven, U., Bender, R., Ziegler, A., & Lange, S. (2007). Vergleich von Messmethoden. *Deutsche Medizinische Wochenschrift, 132*, e69–e73. https://doi.org/10.1055/s-2007-959047

Grözinger, K. E. (2009). *Jüdisches Denken. Theologie – Philosophie – Mystik. Von der Religionskritik der Renaissance zu Orthodoxie und Reform im 19. Jahrhundert* (Band 3). Campus Verlag.

Gründl, M. (2011). *Determinanten physischer Attraktivität – der Einfluss von Durchschnittlichkeit, Symmetrie und sexuellem Dimorphismus auf die Attraktivität von Gesichtern* (Dokumenten-ID 27663) [Habilitationsschrift, Universität Regensburg]. Publikationsserver der Universität Regenburg. https://doi.org/10.5283/epub.27663

Guimond, S., & Roussel, L. (2001). Bragging about one's school grades: Gender stereotyping and students' perception of their abilities in science, mathematics, and language. *Social Psychology of Education, 4*(3–4), 275–293. https://doi.org/10.1023/A:1011332704215

Haag, L. (2018). *Kernkompetenz Klassenführung*. Klinkhardt UTB.

Hackbart, M. (2020). Vorurteile gegenüber queeren Menschen. In M. Hackbart (Hrsg.), *Gesunde Vielfalt pflegen. Zum Umgang mit sexueller und geschlechtlicher Vielfalt in Gesundheit, Pflege und Medizin* (S. 39–54). Waldschlösschen Verlag.

Hackbart, M., Rapior, M., & Thies, B. (2020). Wie werden Erziehungsberatende in Abhängigkeit von Geschlechts- und ethnischer Zugehörigkeit kognitiv repräsentiert? *Zeitschrift für Soziologie der Erziehung und Sozialisation, 40*(2), 116–132. https://doi.org/10.3262/ZSE2002116

Hackbart, M., & Tischoff, C. (2020). Sexuelle und geschlechtliche Vielfalt in der Heteronormativität – eine einleitende Orientierung. In M. Hackbart (Hrsg.), *Gesunde Vielfalt pflegen. Zum Umgang mit sexueller und geschlechtlicher Vielfalt in Gesundheit, Pflege und Medizin* (S. 13–28). Waldschlösschen Verlag.

Haddock, G., & Maio, G. R. (2014). Einstellungen. In K. Jonas, W. Strobe & M. Hewstone (Hrsg.), *Sozialpsychologie* (6. Aufl., S. 197–229). Springer. https://doi.org/10.1007/978-3-642-41091-8_6

Haefele, D. L. (1988). The effects of expectancy on the evaluation of teacher clarity. *Journal of Personnel Evaluation in Education, 2*(3), 251–259. https://doi.org/10.1007/BF00127459

Hagenauer, G., Hascher, T., & Volet, S. E. (2015). Teacher emotions in the classroom: Associations with students' engagement, classroom discipline and the interpersonal teacher-student relationship. *European Journal of Psychology of Education, 30*(4), 385–403. https://doi.org/10.1007/s10212-015-0250-0

Hager, W. (2004). *Testplanung zur statistischen Prüfung psychologischer Hypothesen.* Hogrefe.

Haimerl, C., & Fries, S. (2010). Self-fulfilling prophecies in media-based learning: Content relevance moderates quality expectation effects on academic achievement. *Learning and Instruction, 20*(6), 489–510. https://doi.org/10.1016/j.learninstruc.2009.07.002

Haken, H., & Schiepek, G. (2010). *Synergetik in der Psychologie. Selbstorganisation verstehen und gestalten* (2. Aufl.). Hogrefe.

Hannemann, L. (2020). *Evaluation eines Classroom-Management-Trainings zur Kompetenzförderung im Lehramtsstudium unter Berücksichtigung der Entwicklung und Messung der Lehrkraft-Selbstwirksamkeitserwartung* [Dissertation, Technische Universität Braunschweig]. http://d-nb.info/1216728704

Hansen, E., Zech, N., & Meissner, K. (2017). Placebo und Nocebo. Wie einsetzen bzw. vermeiden? *Der Internist, 58*(10), 1102–1110. https://doi.org/10.1007/s00108-017-0294-0

Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin, 97*(3), 363–386. https://doi.org/10.1037/0033-2909.97.3.363

Hartmann, P., Reuter, M., & Nyborg, H. (2006). The relationship between date of birth and individual differences in personality and general intelligence: A large-scale study. *Personality and Individual Differences, 40*(7), 1349–1362. https://doi.org/10.1016/j.paid.2005.11.017

Harwell, M. (2003). Summarizing Monte Carlo results in methodological research: The single-factor, fixed-effects ANCOVA case. *Journal of Educational and Behavioral Statistics, 28*(2), 45–70. https://doi.org/10.3102%2F10769986028001045

Hattie, J. (2015). *Lernen sichtbar machen* (3. Aufl.). Schneider Verlag Hohengehren.

Heckhausen, J., & Heckhausen, H. (2018). *Motivation und Handeln* (5. Aufl.). Springer. https://doi.org/10.1007/978-3-662-53927-9

Helmke, A. (2017). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (7. Aufl.). Klett Kallmeyer.

Helmke, A., & Helmke, T. (2014). Wie wirksam ist gute Klassenführung? Effiziente Klassenführung ist nicht alles, aber ohne sie geht alles andere gar nicht. *Lernende Schule, 65, 9–12.*

Helmke, A., Rindermann, H., & Schrader, F.-W. (2008). Wirkfaktoren akademischer Leistungen in Schule und Hochschule. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 145–155). Hogrefe.

Hemmerich, W. (2016). *StatistikGuru: Rechner zur Adjustierung des α-Niveaus.* Abgerufen am 20.12.2020 von https://statistikguru.de/rechner/adjustierung-des-alphaniveaus.html

Herrell, J. M. (1971). Galatea in the classroom: Student expectations affect teacher behavior. *Proceedings of the 79th Annual Convention of the American Psychological Association, 6*(Teil 2), 521–522.

Herrmann, U. (Hrsg.). (2019). *Pädagogische Beziehungen. Grundlagen – Praxisformen – Wirkungen.* Beltz Juventa.

Heyder, A., & Kessels, U. (2015). Do teachers equate male and masculine with lower academic engagement? How students' gender enactment triggers gender stereotypes at school. *Social Psychology of Education, 18*(3), 467–485. https://doi.org/10.1007/s11218-015-9303-0

Ho, R., & Mitchell, S. (1982). Students' nonverbal reaction to tutors' warm/cold nonverbal behavior. *Journal of Social Psychology, 118*(1), 121–130. https://doi.org/10.1080/00224545.1982.9924425

Hoefert, H.-W. (2010). *Psychologie in der Arztpraxis.* Hogrefe.

Hofer, M., & Haimerl, C. (2008). Lehrer-Schüler-Interaktion. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 223–232). Hogrefe.

Holtmann, J., Koch, T., Lochner, K., & Eid, M. (2016). A comparison of ML, WLSMV, and Bayesian methods for multilevel structural equation models in small samples: A simulation study. *Multivariate Behavioral Research, 51*(5), 661–680. https://doi.org/10.1080/00273171.2016.1208074

Horstkemper, M. (2000). Geschlecht und Professionalität. Lehrerinnen und Lehrer – Über die Bedeutung der Geschlechterdifferenz. In J. Bastian, W. Helsper, S. Reh & C. Schelle (Hrsg.), *Professionalisierung im Lehrberuf* (S. 87–105). Leske + Budrich.

Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling, 8*(2), 157–174. https://doi.org/10.1207/S15328007SEM0802_1

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica, 64*(2), 157–170. https://doi.org/10.1111/j.1467-9574.2009.00445.x

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huber, F., Meyer, F., & Lenzen, M. (2014). *Grundlagen der Varianzanalyse. Konzeption – Durchführung – Auswertung*. Springer Gabler. https://doi.org/0.1007/978-3-658-05666-7

Hubrig, C., & Herrmann, P. (2014). *Lösungen in der Schule. Systemisches Denken in Unterricht, Beratung und Schulentwicklung*. Carl-Auer.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*(1), 53–69. https://doi.org/10.1037/0033-2909.104.1.53

Ihme, T. A., & Möller, J. (2015). "He who can, does; he who cannot, teaches?": Stereotype threat and preservice teachers. *Journal of Educational Psychology, 107*(1), 300–308. https://doi.org/10.1037/a0037373

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology, 4,* Artikel 386. https://doi.org/10.3389/fpsyg.2013.00386

Jackson, L. A., Hunter, J. E., & Hodge, C. N. (1995). Physical attractiveness and intellectual competence: A meta-analytic review. *Social Psychology Quarterly, 58*(2), 108–122. https://doi.org/10.2307/2787149

Jamieson, D. W., Lydon, J. E., Stewart, G., & Zanna, M. P. (1987). Pygmalion revisited: New evidence for student expectancy effects in the classroom. *Journal of Educational Psychology, 79*(4), 461–466. https://doi.org/10.1037/0022-0663.79.4.461

Jastrow, J. (1901). *Fact and Fable in Psychology*. Macmillan and Co.

Joye, S., & Wilson, J. (2015). Professor age and gender affect student perceptions and grades. *Journal of the Scholarship of Teaching and Learning, 15*(4), 126–138.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology, 9*(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3

Kahraman, B., & Knoblich, G. (2000). „Stechen statt Sprechen": Valenz und Aktivierbarkeit von Stereotypen über Türken. *Zeitschrift für Sozialpsychologie, 31*(1), 31–43. https://doi.org/10.1024/0044-3514.31.1.31

Kaiser, A. (2009). *Ungleiche Bildungschancen schon durch Vornamen? Studie zu Vorurteilen und Vorannahmen von Lehrern.* Abgerufen am 13.06.2019 von https://www.presse.uni-oldenburg.de/mit/2009/390.html

Kelava, A., & Moosbrugger, H. (2020). Deskriptivstatistische Itemanalyse und Testwertbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 143–158). Springer. https://doi.org/10.1007/978-3-662-61532-4_7

Keller, G. (2010). *Disziplinmanagement in der Schulklasse. Unterrichtsstörungen vorbeugen – Unterrichtsstörungen bewältigen* (2. Aufl.). Huber.

Kemper, C. J., Lutz, J., Margraf-Stiksrud, J., Beierlein, C., Kovaleva, A., & Rammstedt, B. (2014). Attraktivitätsrating (AR1). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. https://doi.org/10.6102/zis182

Kessler, T., & Mummendey, A. (2007). Vorurteile und Beziehungen zwischen sozialen Gruppen. In K. Jonas, W. Stroebe und M. Hewstone (Hrsg.), *Sozialpsychologie* (5. Aufl., S. 485–531). Springer. https://doi.org/10.1007/978-3-540-71633-4_14

Kiel, E., Frey, A., & Weiß, S. (2013). *Trainingsbuch Klassenführung*. Klinkhardt UTB.

Kiese-Himmel, C. (2016). *Körperinstrument Stimme. Grundlage, psychologische Bedeutung, Störung*. Springer. https://doi.org/10.1007/978-3-662-49648-0

Kirsch, I. (1985). Response expectancy as a determinant of experience and behavior. *American Psychologist, 40*(11), 1189–1202. https://doi.org/10.1037/0003-066X.40.11.1189

Klauer, K. C. (2020). Soziale Kategorisierung und Stereotypisierung. In L.-E. Petersen & B. Six (Hrsg.), *Stereotype, Vorurteile und soziale Diskriminierung. Theorien, Befunde und Interventionen* (2. Aufl., S. 23–32). Beltz.

Klein, O., Doyen, S., Leys, C., de Saldanha da Gama, P. A. M., Miller, S., Questienne, L., & Cleeremans, A. (2012). Low hopes, high expectations: Expectancy effects and the replicability of behavioral experiments. *Perspectives on Psychological Science, 7*(6), 572–584. https://doi.org/10.1177/1745691612463704

Klein, S. (2011). *Wirksamkeitserwartungen an Einflüsse auf den Wissenserwerb erwachsener Lernender – Herleitung eines Modells für die berufliche Weiterbildung aus der Schulforschung* [Dissertation, Freie Universität Berlin]. Refubium – Freie Universität Berlin Repository. https://doi.org/10.17169/refubium-17190

Kleinke, K., Schlüter, E., & Christ, O. (2017). *Strukturgleichungsmodelle mit Mplus. Eine praktische Einführung* (2. Aufl.). De Gruyter.

Knierim, B., Raufelder, D., & Wettstein, A. (2017). Die Lehrer-Schüler-Beziehung im Spannungsfeld verschiedener Theorieansätze. *Psychologie in Erziehung und Unterricht, 64*(1), 35–48. https://doi.org/10.2378/peu2017.art04d

Ko, S. J., Judd, C. M., & Stapel, D. A. (2009). Stereotyping based on voice in the presence of individuation information: Vocal femininity affects perceived competence but not warmth. *Personality and Social Psychology Bulletin, 35*(2), 198–211. https://doi.org/10.1177%2F0146167208326477

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative-progressive beliefs, and communion. *Journal of Personality and Social Psychology, 110*(5), 675–709. https://doi.org/10.1037/pspa0000046

Kohler, B., & Wacker, A. (2013). Das Angebots-Nutzungs-Modell. Überlegungen zu Chancen und Grenzen des derzeit prominentesten Wirkmodells der Schul- und Unterrichtsforschung. *Die Deutsche Schule, 105*(3), 242–257.

Komarraju, M., Musulkin, S., & Bhattacharya, G. (2010). Role of student-faculty interactions in developing college students' academic self-concept, motivation, and achievement. *Journal of College Student Development, 51*(3), 332–342. https://doi.org/10.1353/csd.0.0137

Koopmans, M., & Stamovlasis, D. (2016). Introducing to education as a complex dynamical system. In M. Koopmans & D. Stamovlasis (Hrsg.), *Complex dynamical systems in education. Concepts, methods and applications* (S. 1–7). Springer Cham. https://doi.org/10.1007/978-3-319-27577-2_1

Kounin, J. S. (2006). *Techniken der Klassenführung.* Waxmann.
Kowai-Bell, N., Guadagno, R. E., Little, T. E., & Ballew, J. L. (2012). Professors are people too: The impact of informal evaluations of professors on students and professors. *Social Psychology of Education, 15*(3), 337–351. https://doi.org/10.1007/s11218-012-9181-7
Kowai-Bell, N., Guadagno, R. E., Little, T., Preiss, N., & Hensley, R. (2011). Rate my expectations: How online evaluations of professors impact students' perceived control. *Computers in Human Behavior, 27*(5), 1862–1867. https://doi.org/10.1016/j.chb.2011.04.009
Krahé, B., & Papakonstantinou, L. (2020). Speaking like a man: Women's pitch as a cue for gender stereotyping. *Sex Roles, 82*(1–2), 94–101. https://doi.org/10.1007/s11199-019-01041-z
Krings, F., & Kluge, A. (2020). Altersvorurteile. In L.-E. Petersen & B. Six (Hrsg.), *Stereotype, Vorurteile und soziale Diskriminierung. Theorien, Befunde und Interventionen* (2. Aufl., S. 153–161). Beltz.
Kriz, J. (2007). *Grundkonzepte der Psychotherapie* (6. Aufl.). Beltz.
Kronschläger, T. (2020). Geschlechtssensible Sprache. In M. Hackbart (Hrsg.), *Gesunde Vielfalt pflegen. Zum Umgang mit sexueller und geschlechtlicher Vielfalt in Gesundheit, Pflege und Medizin* (S. 29–37). Waldschlösschen Verlag.
Kruglanski, A. W., Jasko, K., Milyavsky, M., Chernikova, M., Webber, D., Pierro, A., & di Santo, D. (2018). Cognitive consistency theory in social psychology: A paradigm reconsidered. *Psychological Inquiry, 29*(2), 45–59. https://doi.org/10.1080/1047840X.2018.1480619
Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Waxmann.
Lambert, M. J., & Kleinstäuber, M. (2016). Wenn sich Menschen ändern – Veränderungsprozesse in der Psychotherapie und ihre Beziehung zu spezifischen Therapietechniken und allgemeinen Wirkfaktoren. *Verhaltenstherapie, 26*(1), 32–39. https://doi.org/10.1159/000442372
Landy, D., & Sigall, H. (1974). Beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology, 29*(3), 299–304. https://doi.org/10.1037/h0036018
Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*(3), 390–423. https://doi.org/10.1037/0033-2909.126.3.390
Laux, L. (2008). *Persönlichkeitspsychologie* (2. Aufl.). Kohlhammer.
Leventhal, L., Perry, R. P., & Abrami, P. C. (1977). Effects of lecturer quality and student perception of lecturer's experience on teacher ratings and student achievement. *Journal of Educational Psychology, 69*(4), 360–374. https://doi.org/10.1037/0022-0663.69.4.360
Lewandowski, G. W., Higgins, E., & Nardonne, N. N. (2012). Just a harmless website?: An experimental examination of ratemyprofessor.com's effect on student evaluations. *Assessment & Evaluation in Higher Education, 37*(8), 987–1002. https://doi.org/10.1080/02602938.2011.594497
Lewin, K., Lippit, R., & White, R. K. (1939). Patterns of aggressive behavior in experimentally created "social climates". *Journal of Social Psychology, 10*(2), 271–299. https://doi.org/10.1080/00224545.1939.9713366

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: A review. *Educational Research, 41*(1), 63–76. https://doi.org/10.1080/0013188990410106

Lies, J., Vaih-Baur, & Verčič, D. (2015). PR-Theorien. In J. Lies (Hrsg.), *Theorien des PR-Managements. Geschichte – Basiswissenschaften – Wirkungsdimensionen* (S. 169–172). Springer Gabler. https://doi.org/10.1007/978-3-658-06997-1_6

Lietz, P. (2006). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation, 32*(4), 317–344. https://doi.org/10.1016/j.stueduc.2006.10.002

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*(6), 1123–1135. https://doi.org/10.1037/a0021276

Liu, J., Hu, J., & Furutan, O. (2013). The influence of student perceived professors' "hotness" on expertise, motivation, learning outcomes, and course satisfaction. *Journal of Education in Business, 88*(2), 94–100. https://doi.org/10.1080/08832323.2011.652695

Locher, C., Hasler, S., & Gaab, J. (2016). Placebos in der Psychotherapieforschung – Eine systematische Analyse am Beispiel der systematischen Desensibilisierung. *Verhaltenstherapie, 26*(1), 9–20. https://doi.org/10.1159/000443464

Lorenz, G. (2019). *Selbsterfüllende Prophezeiungen in der Schule. Leistungserwartungen von Lehrkräften und Kompetenzen von Kindern mit Zuwanderungshintergrund*. Springer VS. https://doi.org/10.1007/978-3-658-19881-7

Lorenz, G., Gentrup, S., Kristen, C., Stanat, P., & Kogan, I. (2016). Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 68*(1), 89–111. https://doi.org/10.1007/s11577-015-0352-3

Lühmann, H. (2019). Freuds Konzept der Übertragung – Eine unverzichtbare pädagogische Beziehung in der Schule. In U. Herrmann (Hrsg.), *Pädagogische Beziehungen. Grundlagen – Praxisformen – Wirkungen* (S. 42–60). Beltz Juventa.

Ludwig, P. H. (2018). Erwartungseffekt. In D. H. Rost, J. Sparfeldt & S. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (5. Aufl., S. 141–147). Beltz.

Ma, C., Panaccione, N. R., Nguyen, T. M., Guizetti, L., Parker, C. E., Hussein, I. M., Vande Casteele, N., Khanna, R., Dulai, P. S., Singh, S., Feagan, B. G., & Jairath, V. (2019). Adverse events and nocebo effects in inflammatory bowel disease: A systematic review and meta-analysis of randomized controlled trials. *Journal of Crohn's and Colitis, 13*(9), 1201–1216. https://doi.org/10.1093/ecco-jcc/jjz087

Machunsky, M. (2020). Substereotypisierung. In L.-E. Petersen & B. Six (Hrsg.), *Stereotype, Vorurteile und soziale Diskriminierung. Theorien, Befunde und Interventionen* (2. Aufl., S. 45–53). Beltz.

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*(4), 291–303. https://doi.org/10.1007/s10755-014-9313-4

Maier, M. A., Elliot, A. J., Lee, B., Lichtenfeld, S., Barchfeld, P., & Pekrun, R. (2013). The influence of red on impression formation in a job application context. *Motivation and Emotion, 37*(3), 389–401. https://doi.org/10.1007/s11031-012-9326-1

Mancus, D. S. (1992). Influence of male teachers on elementary school children's stereotyping to teacher competence. *Sex Roles, 26*(3–4), 109–128. https://doi.org/10.1007/BF00289753

Sánchez, M. M., Martinez-Pecino, R., Rodríguez, Y. T., & Melero, P. T. (2011). Student perspectives on the university professor role. *Social Behavior and Personality: An international journal, 39*(4), 491–496. https://doi.org/10.2224/sbp.2011.39.4.491

Martin, K. J., & Smith, L. R. (1990). *Effect of teacher age and gender on student perception* (ED347162). Educational Resources Information Centre (ERIC), U.S.A.

Marzano, R. J., Marzano, J. S., & Pickering, D. J. (2003). *Classroom management that works: Research-based strategies for every teacher*. Association for Supervision and Curriculum Development.

Mehrabian, A. (1969). Some referents and measures of nonverbal behavior. *Behavior Research Methods & Instrumentation, 1*(6), 203–207. https://doi.org/10.3758/BF03208096

Meltzer, A. L., & McNulty, J. K. (2011). Contrast effects of stereotypes: "Nurturing" male professors are evaluated more positively than "nurturing" female professors. *The Journal of Men's Studies, 19*(1), 57–64. https://doi.org/10.3149%2Fjms.1901.57

Merzyn, G. (2017). Merkmale guter Lehrer in Physik, Chemie, Biologie. Ein Überblick. *Physik und Didaktik in Schule und Hochschule, 16*(1), 67–80.

Micari, M., & Pazos, P. (2012). Connecting to the professor: Impact of the student-faculty relationship in a highly challenging course. *College Teaching, 60*(2), 41–47. https://doi.org/10.1080/87567555.2011.627576

Michael, R. B., Garry, M., & Kirsch, I. (2012). Suggestion, cognition, and behavior. *Current Directions in Psychological Science, 21*(3), 151–156. https://doi.org/10.1177%2F0963721412446369

Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The development of children's gender-science-stereotypes: A meta-analysis of 5 decades of U.S. draw-a-scientist studies. *Child Development, 89*(6), 1943–1955. https://doi.org/10.1111/cdev.13039

Misamer, M. (2019). *Macht und Machtmittel in der Schule. Eine empirische Untersuchung*. Jacobs Verlag. http://melaniemisamer.de/

Morris, T. L., Gorham, J., Cohen, S. H., & Huffmann, D. (1996). Fashion in the classroom: Effects of attire on student perceptions of instructors in college classes. *Communication Education, 45*(2), 135–148. https://doi.org/10.1080/03634529609379043

Mottet, T. P., Beebe, S. A., Raffeld, P. C., & Medlock, A. L. (2004). The effects of student verbal and nonverbal responsiveness on teacher self-efficacy and job satisfaction. *Communication Education, 53*(2), 150–163. https://doi.org/10.1080/03634520410001682410

Müller, B. C. N., Oude Groote Beverborg, A., & Glock, S. (2017). Pre-service teachers' academic judgments of overweight students. *Social Psychology of Education, 20*(4), 897–913. https://doi.org/10.1007/s11218-017-9400-3.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30

Münzer, S., & Brünken, R. (2018). Aptitude-Treatment-Interaktion. In D. H. Rost, J. R. Sparfeldt & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (5. Aufl., S. 14–20). Beltz.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus. Statistical analysis with latent variables. User's Guide* (8. Aufl.). Muthén & Muthén. https://www.statmodel.com/html_ug.shtml

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin, 83*(4), 602–627. https://doi.org/10.1037/0033-2909.83.4.602

Nadler, J. T., Berry, S. A., & Stockdale, M. S. (2013). Familiarity and sex based stereotypes on instant impressions of male and female faculty. *Social Psychology of Education, 16*(3), 517–539. https://doi.org/10.1007/s11218-013-9217-7

Nickel, H. (1985). Die Lehrer-Schüler-Beziehung aus der Sicht neuerer Forschungsergebnisse. Ein transaktionales Modell (1976). In R. Biermann (Hrsg.), *Interaktion-Schule-Unterricht* (S. 254–280). Wissenschaftliche Buchgesellschaft.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R. W., … Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences, U.S.A., 106*(26), 10593–10597. https://doi.org/10.1073/pnas.0809921106

Ntemiris, N. (2011). *Gouvernementalität und Kindheit. Transformationen generationaler Ordnung in Diskursen und in der Praxis.* Springer VS. https://doi.org/10.1007/978-3-531-92827-2

Oleszkiewicz, A., Pisanski, K., Lachowicz-Tabaczek, K., & Sorokowska, A. (2017). Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. *Psychonomic Bulletin & Review, 24*(3), 856–862. https://doi.org/10.3758/s13423-016-1146-y

Olshansky, B. (2007). Placebo and nocebo in cardiovascular health. Implications for healthcare, research, and the doctor-patient relationship. *Journal of the American College of Cardiology, 49*(4), 415–421. https://doi.org/10.1016/j.jacc.2006.09.036

Ovid (2003). Metamorphoses/Verwandlungen. In A. Aurnhammer & D. Martin (Hrsg.*), Mythos Pygmalion. Texte von Ovid bis John Updike* (S. 10–13). Reclam.

Pawlik, K., & Buse, L. (1979). Self-attribution as a differential, psychological moderator variable: Verification and clarification of Eysenck's astrology-personality correlations. *Zeitschrift für Sozialpsychologie, 10*(1), 54–69.

Pendry, L. (2014). Soziale Kognition. In K. Jonas, W. Stroebe & M. Hewstone (Hrsg.), *Sozialpsychologie* (6. Aufl., S. 107–140). Springer. https://doi.org/10.1007/978-3-642-41091-8_4

Pennings, H. J. M., & Mainhard, T. (2016). Analyzing teacher-student interactions with state space grids. In M. Koopmans & D. Stamovlasis (Hrsg.), *Complex dynamical systems in education. Concepts, methods and applications* (S. 233–271). Springer Cham. https://doi.org/10.1007/978-3-319-27577-2_12

Perrez, M., Huber, G. L., & Geißler, K. A. (2006). Psychologie der pädagogischen Interaktion. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (5. Aufl., S. 357–421). Beltz.

Perry, R. P., Abrami, P. C., Leventhal, L., & Check, J. (1979). Instructor reputation: An expectancy relationship involving student ratings and achievement. *Journal of Educational Psychology, 71*(6), 776–787. https://doi.org/10.1037/0022-0663.71.6.776

Petersen, L.-E., & Six, B. (Hrsg.). (2020). *Stereotype, Vorurteile und soziale Diskriminierung. Theorien, Befunde und Interventionen* (2. Aufl.). Beltz.

Peterson, E. R., Rubie-Davies, C., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction, 42*, 123–140. https://doi.org/10.1016/j.learninstruc.2016.01.010

Plante, I., Théorêt, M., & Favreau, O. E. (2009). Student gender stereotypes: Contrasting the perceived maleness and femaleness of mathematics and language. *Educational Psychology, 29*(4), 385–405. https://doi.org/10.1080/01443410902971500

Quin, D. (2016). Longitudinal and contextual associations between teacher-student relationships and student engagement: A systematic review. *Review of Educational Research, 87*(2), 345–387. https://doi.org/10.3102%2F0034654316669434

Raabe, T., & Beelmann, A. (2009). Entwicklungspsychologische Grundlagen. In A. Beelmann & K. J. Jonas (Hrsg.), *Diskriminierung und Toleranz. Psychologische Grundlagen und Anwendungsperspektiven* (S. 113–135). VS Verlag. https://doi.org/10.1007/978-3-531-91621-7_6

Rahm, T., Heise, E., & Schuldt, M. (2017). Measuring the frequency of emotions: Validation of the Scale of Positive and Negative Experience (SPANE) in Germany. *PLoS ONE, 12*(2), Artikel e0171288. https://doi.org/10.1371/journal.pone.0171288

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76*(1), 85–97. https://doi.org/10.1037/0022-0663.76.1.85

Reavis, C. A. (1979). A study of the effects of prefatory remarks on teacher evaluation. *Journal of Educational Research, 72*(3), 173–177. https://doi.org/10.1080/00220671.1979.10885146

Reber, J. S., Ridge, R. D., & Downs, S. D. (2017). Perceptual and behavioral effects of expectations formed by exposure to positive or negative ratemyprofessors.com evaluations. *Cogent Psychology, 4*(1), Artikel 1338324. https://doi.org/10.1080/23311908.2017.1338324

Reusser, K., & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität – Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht: Einleitung und Überblick. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 9–32). Waxmann.

Reynolds, D. (2007). Restraining Golem and harnessing Pygmalion in the classroom: A laboratory study of managerial expectations and task design. *Academy of Management Learning & Education, 6*(4), 475–483. https://doi.org/10.5465/AMLE.2007.27694947

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Die Erfassung des Flow-Erlebens. In J. Stiensmeier-Pelster & F. Rheinberg (Hrsg.), *Diagnostik von Motivation und Selbstkonzept* (S. 261–279). Hogrefe.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2019). FKS. Flow-Kurzskala [Verfahrensdokumentation, Fragebogen und Normtabelle]. In Leibniz-Institut für Psychologie (ZPID) (Hrsg.), *Open Test Archive*. Leibniz-Institut für Psychologie. https://doi.org/10.23668/psycharchives.4488

Richey, P. (2016). *Lehrer-Schüler-Beziehung. Eine empirische Studie zu normativen Lehrer- und Schülererwartungen aus Lehrer-, Schüler- und Beobachterperspektive*. Schneider Verlag Hohengehren.

Riniolo, T. C., Johnson, K. C., Sherman, T. R., & Misso, J. A. (2006). Hot or not: Do professors perceived as physically attractive receive higher student evaluations? *Journal of General Psychology, 133*(1), 19–35. https://doi.org/10.3200/GENP.133.1.19-35

Ritts, V., Patterson, M. L., & Tubbs, M. E. (1992). Expectations, impressions, and judgements of physically attractive students: A review. *Review of Educational Research, 62*(4), 413–426. https://doi.org/10.3102%2F00346543062004413

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review, 9*(1), 32–47. https://doi.org/10.1207%2Fs15327957pspr0901_3

Rodríguez-García, J.-M., & Wagner, U. (2009). Learning to be prejudiced: A test of unidirectional and bidirectional models of parent-offspring socialization. *International Journal of Intercultural Relations, 33*(6), 516–523. https://doi.org/10.1016/j.ijintrel.2009.08.001

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*(2), 95–103. https://doi.org/10.1037/h0045357

Rogers, C. R. (1981). *Der neue Mensch.* Klett-Cotta.

Romano, S. T., & Bordieri, J. E. (1989). Physical attractiveness stereotypes and students' perceptions of college professors. *Psychological Reports, 64*(3, Teil 2), 1099–1102. https://doi.org/10.2466/pr0.1989.64.3c.1099

Rosemann, B. (1978). Bedingungsvariablen der Lehrer-Schüler-Beziehung. Erwartungskonkordanz und das Verhalten von Lehrern und Schülern. *Psychologie in Erziehung und Unterricht, 25*(1), 39–49.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*(4), 283–294. https://doi.org/10.1037/h0026086

Rosenthal, R. (1966). *Experimenter effects in behavioral research.* Appleton-Century-Crofts.

Rosenthal, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current Directions in Psychological Science,3*(6), 176–179. https://doi.org/10.1111/1467-8721.ep10770698

Rosenthal, R. (2009). Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal & R. L. Rosnow (Hrsg.), *Artifacts in behavioral research* (S. 138–210). Oxford University Press.

Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science, 8*(3), 183–189. https://doi.org/10.1002/bs.3830080302

Rosenthal, R., & Jacobson, L. (1966). Teachers' expectancies: Determinants of pupils' IQ gains. *Psychological Reports, 19*(1), 115–118. https://doi.org/10.2466/pr0.1966.19.1.115

Rosenthal, R., & Jacobson, L. (1971). *Pygmalion im Unterricht. Lehrererwartungen und Intelligenzentwicklung der Schüler.* Beltz.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences, 1*(3), 377–386. https://doi.org/10.1017/S0140525X00075506

Rubie-Davis, C. M. (2007). Classroom interactions: Exploring the practices of high- and low-expectation teachers. *British Journal of Educational Psychology, 77*(2), 289–306. https://doi.org/10.1348/000709906X101601

Rudolph, U. (2009). Erwartung und Anreiz. In V. Brandstätter & J. H. Otto (Hrsg.), *Handbuch Allgemeine Psychologie – Motivation und Emotion* (S. 21–28). Hogrefe.

Saretsky, G. (1972). The OEO P.C. experiment and the John Henry effect. *Phi Delta Kappan, 53*(9), 579–581.

Schallnus, R. N. (2006). *Mitarbeiterqualifizierung und Wissensnutzung in Konzernen und Unternehmungsnetzwerken* [Dissertation, Freie Universität Berlin]. Refubium – Freie Universität Berlin Repository. https://doi.org/10.17169/refubium-17190

Schaub, H.-A. (2013). *Die pädagogische Beziehung zwischen Lehrern und Schülern. Seminar auf der Eröffnungstagung 06.-07.09.2013 Bad Mergentheim.* Abgerufen am 23.05.2014 von http://www.ptz.de/downloadbereich.66.0.html

Schiepek, G., Eckert, H., & Kravanja, B. (2013). *Grundlagen systemischer Therapie und Beratung. Psychotherapie als Förderung von Selbstorganisationsprozessen.* Hogrefe.

Schneewind, K. A., & Böhmert, B. (2009a). *Grundschulkinder kompetent erziehen. Der interaktive Elterncoach „Freiheit in Grenzen“* (2. Aufl.). Huber.

Schneewind, K. A., & Böhmert, B. (2009b). *Jugendliche kompetent erziehen. Der interaktive Elterncoach „Freiheit in Grenzen“*. Huber.

Schneewind, K. A., & Böhmert, B. (2009c). *Vorschulkinder kompetent erziehen. Der interaktive Elterncoach „Freiheit in Grenzen“*. Huber

Schweer, M. K. W. (Hrsg.). (2017). *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge.* Springer VS. https://doi.org/10.1007/978-3-658-15083-9

Schweer, M. K. W. (2019). Soziale Wahrnehmung im Unterricht – Grundlage pädagogischer Beziehungen. In U. Herrmann (Hrsg.), *Pädagogische Beziehungen. Grundlagen – Praxisformen – Wirkungen* (S. 61–72). Beltz Juventa.

Schweer, M. K. W., & Rosemann, B. (1995). Qualität der Lehre. Bedingungsvariablen des studentischen Urteils. *Zeitschrift für Pädagogische Psychologie, 9*(3–4), 189–196.

Seidel, T. (2014). Angebots-Nutzungs-Modell in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma. *Zeitschrift für Pädagogik, 60*(6), 850–866. https://doi.org/10.3262/ZP1406850

Shapiro, J. R., King, E. B., & Quiñones, M. A. (2007). Expectations of obese trainees: How stigmatized trainee characteristics influence training effectiveness. *Journal of Applied Psychology, 92*(1), 239–249. https://doi.org/10.1037/0021-9010.92.1.239

Siebertz-Reckzeh, K., & Hofmann, H. (2017). Sozialisationsinstanz Schule. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (3. Aufl., S. 3–26). Springer VS. https://doi.org/10.1007/978-3-658-15083-9_1

Smith, M. L. (1980). Teacher expectations. *Evaluation in Education, 4*, 53–55. https://doi.org/10.1016/0191-765X(80)90015-4

Smyth, F. L., & Nosek, B. A. (2015). On the gender-science stereotypes held by scientists: Explicit accord with gender-ratios, implicit accord with scientific identity. *Frontiers in Psychology, 6*, Artikel 415. https://doi.org/10.3389/fpsyg.2015.00415

Sohr-Preston, S. L., Boswell, S. S., McCaleb, K., & Robertson, D. (2016). Professor gender, age, and "hotness" in influencing college students' generation and interpretation of professor ratings. *Higher Learning Research Communications, 6*(3). https://doi.org/10.18870/hlrc.v6i3.328

Sorhagen, N. S. (2013). Early teacher expectations disproportionately affect poor children's high school performance. *Journal of Educational Psychology, 105*(2), 465–477. https://doi.org/10.1037/a0031754

Spears, R., & Tausch, N. (2014). Vorurteile und Intergruppenbeziehungen. In K. Jonas, W. Stroebe & M. Hewstone (Hrsg.), *Sozialpsychologie* (6. Aufl., S. 507–564). Springer. https://doi.org/10.1007/978-3-642-41091-8_14

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240–261. https://doi.org/10.1037/0033-2909.124.2.240

Stangor, C. (2009). The study of stereotyping, prejudice, and discrimination within social psychology. A quick history of theory and research. In T. D. Nelson (Hrsg.), *Handbook of prejudice, stereotyping, and discrimination* (S. 1–22). Psychology Press.

Statistisches Bundesamt (2020a). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2019*. Statistisches Bundesamt. https://www.destatis.de/DE/Service/Bibliothek/_publikationen-fachserienliste-1.html

Statistisches Bundesamt (2020b). *Bildung und Kultur. Personal an Hochschulen.* Abgerufen am 02.01.2021 von https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/_inhalt.html#sprg234556

Statistisches Bundesamt (2020c). *Bildung und Kultur. Studierende an Hochschulen.* Abgerufen am 02.01.2021 von https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/_inhalt.html#sprg234556

Statistisches Bundesamt (2020d, 20. Oktober). *Lehrkräfte in allgemeinbildenden und beruflichen Schulen insgesamt sowie Anteil der weiblichen Lehrkräfte.* Abgerufen am 02.01.2021 von https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Schulen/Tabellen/allgemeinbildende-beruflicheschulen-lehrkraefte.html

Steins, G. (2014). *Sozialpsychologie des Schulalltags. Grundlagen und Anwendungen* (2. Aufl.). Pabst.

Strunk, G., & Schiepek, G. (2006). *Systemische Psychologie. Eine Einführung in die komplexen Grundlagen menschlichen Verhaltens*. Elsevier Spektrum.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin, 135*(6), 859–884. https://doi.org/10.1037/a0017364

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgements of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743–762. https://doi.org/10.1037/a0027627

Sudman, S., Rath, D., Forkmann, T., Scherer, A., & Gauggel, S. (2014). *Lern- und Leistungsmotivation im Studium (LLMS).* Psychometrikon. https://doi.org/10.6099/1000138

Tauber, R. T. (1997). *Self-fulfilling prophecy. A practical guide to its use in education.* Praeger.

Tausch, R. (2017). Personzentriertes Verhalten von Lehrern in Unterricht und Erziehung. In. M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (3. Aufl., S. 191–212). Springer VS. https://doi.org/10.1007/978-3-658-15083-9_8

Tenenbaum, H. R., & Leaper, C. (2002). Are parents' gender schemas related to their children's gender-related cognitions? A meta-analysis. *Developmental Psychology, 38*(4), 615–630. https://doi.org/10.1037/0012-1649.38.4.615

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*(2), 253–273. https://doi.org/10.1037/0022-0663.99.2.253

Terry, R. L., & McIntosh, D. E. (1988). Do students' expectancies affect their course evaluations? *Educational and Psychological Measurement, 48*(3), 787–798. https://doi.org/10.1177%2F0013164488483028

Thielsch, M. T., & Hirschfeld, G. (2014a). Münsteraner Fragebogen zur Evaluation von Seminaren – revidiert (MFE-Sr). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS).* https://doi.org/10.6102/zis86

Thielsch, M. T., & Hirschfeld, G. (2014b). Münsteraner Fragebogen zur Evaluation von Vorlesungen – revidiert (MFE-Vr). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS).* https://doi.org/10.6102/zis85

Thies, B. (2010). *Kognitive Repräsentationen in der Grundschule. Befunde zur Interaktionsregulation im Unterrichtsalltag*. Lang. https://doi.org/10.3726/978-3-653-00316-1

Thies B. (2017). Forschungszugänge zur Lehrer-Schüler-Interaktion. Ein historischer Abriss. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (3. Aufl., S. 65–88). Springer VS. https://doi.org/10.1007/978-3-658-15083-9_3

Thies, B., Misamer, M., Aurin, S. S., & Hackbart, M. (2016). Transparenz in Schule und Unterricht: Pädagogisch-psychologische Zugänge und Handlungsempfehlungen. In K. Moegling & S. Schude (Hrsg.), *Transparenz im Unterricht und in der Schule. Teil 1: Theorie und Praxis transparenten Unterrichts und transparenter Schulorganisation* (S. 101–126). Prolog-Verlag.

Thompson, A. G. H., & Suñol, R. (1995). Expectations as determinants of patient satisfaction: Concepts, theory and evidence. *International Journal for Quality in Health Care, 7*(2), 127–141. https://doi.org/10.1093/intqhc/7.2.127

Tiedemann, J. (1995). Geschlechtstypische Erwartungen von Lehrkräften im Mathematikunterricht der Grundschule. *Zeitschrift für Pädagogische Psychologie, 9*(3–4), 153–161.

Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics, 50*(1), 49–62. https://doi.org/10.1023/A:1020518104346

Timmermanns, S., & Böhm, M. (Hrsg.). (2020). *Sexuelle und geschlechtliche Vielfalt. Interdisziplinäre Perspektiven aus Wissenschaft und Praxis.* Beltz Juventa.

Tobisch, A., & Dresel, M. (2017). Negatively or positively biased? Dependencies of teachers' judgments and expectations based on students' ethnic and social backgrounds. *Social Psychology of Education,* 20(4), 731–752. https://doi.org/10.1007/s11218-017-9392-z

Tollefson, N., & Wigington, H. (1986). Teacher-generated and student-generated variability in teacher effectiveness ratings. *Instructional Science, 15*, 109–120. https://doi.org/10.1007/BF00139607

Towler, A., & Dipboye, R. L. (2006). Effects of trainer reputation and trainees' need for cognition on training outcomes. *Journal of Psychology, 140*(6), 549–564. https://doi.org/10.3200/JRLP.140.6.549-564

Tröster, H., & Pulz, I. (2020). Stigma und Stigmabewältigung. In L.-E. Petersen & B. Six (Hrsg.), *Stereotype, Vorurteile und soziale Diskriminierung. Theorien, Befunde und Interventionen* (2. Aufl., S. 173–184). Beltz.

Tschacher, W. (1997). *Prozessgestalten.* Hogrefe.

Urban, D., & Mayerl, J. (2018). *Angewandte Regressionsanalyse: Theorie, Technik und Praxis* (5. Aufl.). Springer VS. https://doi.org/10.1007/978-3-658-01915-0

Von Foerster, H. (1997). Abbau und Aufbau. In F. B. Simon (Hrsg.), *Lebende Systeme. Wirklichkeitskonstruktionen in der Systemischen Therapie* (S. 32–51). Suhrkamp.
Voss, R., & Eicher, R. (2015). Erwartungen an Studierende als Passung für die Lehre. Einblicke in die mentalen Muster von Hochschullehrern an Schweizer Fachhochschulen. *Das Hochschulwesen, 63*(3–4), 132–140.
Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204. https://doi.org/10.1037/a0036620
Watzlawick, P., Beavin, J. H., & Jackson, D. D. (2017). *Menschliche Kommunikation. Formen, Störungen, Paradoxien* (13. Aufl.). Hogrefe.
Weber, H. (2012). Methoden der Persönlichkeitsforschung. In H. Weber & T. Rammsayer (Hrsg.), *Differentielle Psychologie – Persönlichkeitsforschung* (S. 21–44). Hogrefe.
Wegner, C., Dück, A., & Grotjohann, N. (2013). Emotion und Interesse als Grundlage für nachhaltiges Lernen begabter Schüler? Eine empirische Studie in der sechsten Jahrgangsstufe von Gymnasien. *Journal für Didaktik und Biowissenschaften, 4*, 42–56.
Weiber, R., & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2. Aufl.). Springer Gabler. https://doi.org/10.1007/978-3-642-35012-2
Wettstein, A., Ramseier, E., Scherzinger, M., & Gasser, L. (2016). Unterrichtstörungen aus Lehrer- und Schülersicht. Aggressive und nicht aggressive Störungen im Unterricht aus Sicht der Klassen., einer Lehrperson und der Schülerinnen und Schüler. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 48*(4), 171–183. https://doi.org/10.1026/0049-8637/a000159
Wettstein, A., Scherzinger, M., & Wyler, S. (2016). Fragebogen zur Erfassung von Störungen im Unterricht: die Faktorenstruktur der Schülerinnen- und Schülerversion. *Empirische Sonderpädagogik, 8*(2), 189–202.
Werth, L., Denzler, M., & Mayer, J. (2020). *Sozialpsychologie – Das Individuum im sozialen Kontext. Wahrnehmen – Denken – Fühlen* (2. Aufl.). Springer. https://doi.org/10.1007/978-3-662-53897-5
Werth, L., Seibt, B., & Mayer, J. (2020). *Sozialpsychologie – Der Mensch in sozialen Beziehungen. Interpersonale und Intergruppenprozesse* (2. Aufl.). Springer. https://doi.org/10.1007/978-3-662-53899-9
Wickström, G., & Bendix, T. (2000). The „Hawthorne effect“ – what did the original Hawthorne studies actually show? *Scandinavian Journal of Work, Environment & Health, 26*(4), 363–367. https://doi.org/10.5271/sjweh.555
Wigington, H., Tollefson, N., & Rodriguez, E. (1989). Students' ratings of instructors revisited: Interactions among class and instructor variables. *Research in Higher Education, 30*(3), 331–344. https://doi.org/10.1007/BF00992608
Williams, T. I., Rose, R., & Chisholm, S. (2006). What is the function of nail biting: An analog assessment study. *Behaviour Research and Therapy, 45*(5), 989–995. https://doi.org/10.1016/j.brat.2006.07.013
Witte, E. H. (2017). *Systemtheorie und Sozialpsychologie* [working paper]. https://doi.org/10.13140/RG.2.2.21827.40480
Wolfrath, U. (2020). Implizite Persönlichkeitstheorien. In L.-E. Petersen & B. Six (Hrsg.), *Stereotype, Vorurteile und soziale Diskriminierung. Theorien, Befunde und Interventionen* (2. Aufl., S. 73–81). Beltz.

Wunder, E. (2003). Self-attribution, sun-sign traits, and the alleged role of favourableness as a moderator variable: Long-term effect or artefact? *Personality and Individual Differences, 35*(8), 1783–1789. https://doi.org/10.1016/S0191-8869(03)00002-3

Yermack, J., & Forsyth, D. R. (2016). Students' implicit theories of university professors. *Scholarship of Teaching and Learning in Psychology, 2*(3), 169–178. https://doi.org/10.1037/stl0000067

Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* [Dissertation, University of California]. Abgerufen am 12.01.2020 von https://www.statmodel.com/download/Yudissertation.pdf

Zander, L., Wolter, I., Latsch, M., & Hannover, B. (2015). Qualified for teaching physics? How perspective teachers perceive teachers with a migration background – and how it's really about "him" or "her". *International Journal of Gender, Science and Technology, 7*(2), 255–279.

Zhang, W., Robertson, J., Jones, A. C., Dieppe, P. A., & Doherty, M. (2008). The placebo effect and its determinants in osteoarthritis: Meta-analysis of randomised controlled trials. *Annals of the Rheumatic Diseases, 67*(12), 1716–1723. https://doi.org/10.1136/ard.2008.092015